```
GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on:            February 11, 2003, 11:46:34 ; Search time 37.7736 Seconds
                                                 (without alignments)
                                                 1605.064 Million cell updates/sec

Title:             US-10-021-121-2
Perfect score:     2450
Sequence:          1 MGPPHSGPGGVRVGALILLG.........TTLLRQRASVEABAQCHGPL 455

Scoring table:     Gapop 10.0 , Gapext 0.5
Searched:          908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters:     908470

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 50 summaries

Database :  A.Geneseq 101002:*
    1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
    2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
    3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
    4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
    5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
    6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
    7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
    8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
    9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
   10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
   11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
   12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
   13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
   14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
   15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
   16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
   17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
   18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
   19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
   20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
   21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
   22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
   23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2450	100.0	455	18	AAW33698	AL-2-long (AL-2l)
2	1841	75.1	340	18	AAW33699	AL-2-short (AL-2s)
3	1841	75.1	340	18	AAW31544	Human cytokine Ler
4	1841	75.1	340	18	AAW16637	NLERK2 ligand for
5	1835	74.9	340	18	AAW17081	EPH family ligand
6	1827	74.6	340	18	AAW46615	Human transmembran
7	628.5	25.7	334	17	AAW00287	Mouse Eph receptor
8	628.5	25.7	336	17	AAR22742	Murine hepatoma tr
9	623	25.4	346	15	AAR55059	Elk tyrosine kinas
10	623	25.4	346	17	AAR91930	Human cytokine elk
11	623	25.4	346	18	AAW36055	Human elk-L protei
12	623	25.4	346	18	AAW19249	Human elk ligand p
13	623	25.4	346	19	AAW44323	Human elk-L, Homo
14	622	25.4	346	17	AAR82606	Eph transmembrane
15	621.5	25.4	331	17	AAW00288	Chicken Eph recept
16	620.5	25.3	333	17	AAW06337	Full length ligand
17	620.5	25.3	333	17	AAR94655	Ligand for recepto
18	620.5	25.3	333	17	AAR92743	Human hepatoma tra
19	620.5	25.3	333	17	AAR89287	Human LERK-5 cytok
20	610.5	24.9	308	17	AAW11308	Receptor-type tyro
21	610.5	24.9	308	17	AAW06334	Ligand #2 for rece
22	610.5	24.9	308	17	AAR94656	Ligand for recepto
23	483	19.7	89	21	AAY71438	Human ephrin B3-C-
24	458.5	18.7	308	21	AAY96781	Ephrin-B1-Ephrin-B
25	454	18.5	658	21	AAY96782	Ephrin-B2-Ephrin-B
26	447	18.2	229	21	AAE24019	Murine ephrin B2 1
27	443	18.1	229	23	AAE24020	Murine ephrin B2 1
28	431.5	17.6	195	17	AAW06333	Ligand #1 for rece
29	431.5	17.6	195	18	AAW11307	Receptor-type tyro
30	284.5	11.6	92	22	AAM37671	Peptide #1708 enc
31	284.5	11.6	92	23	ABG46524	Human peptide enco
32	196.5	8.0	136	22	AAM37534	Peptide #11571 enc
33	196.5	8.0	136	23	ABG46394	Human peptide enco
34	195	8.0	136	23	AAY71437	Human ephrin B2-C-
35	193.5	7.9	106	21	AAY71436	Human ephrin B2 C-
36	192	7.8	82	21	AAB54187	Human pancreatic c
37	192	7.8	652	22	AAB58621	Drosophila melanog
38	192	7.8	652	22	AAB65548	Drosophila melanog
39	181	7.4	234	16	AAR82605	Eph transmembrane
40	179	7.3	238	16	AAR71481	Human hek-L protei
41	176	7.2	184	17	AAW02587	Lerk-6 protein. M
42	176	7.2	184	17	AAR91283	LERK-6 protein. M
43	176	7.2	184	20	AAY06820	Murine LERK-6 poly
44	176	7.2	209	17	AAX06820	Mouse EPH receptor
45	176	7.2	209	19	AAW71006	Amino acid sequenc
46	175.5	7.2	213	20	AAX06822	Human EPH receptor
47	172.5	7.0	218	21	AAS54058	Human pancreatic c
48	172	7.0	200	17	AAR94767	Chicken EPH recept
49	172	7.0	200	19	AAW71007	Amino acid sequenc
50	169.5	6.9	201	16	AAR71482	Human hek-L protei

ALIGNMENTS

```
RESULT 1
AAW33698

ID   AAW33698 standard; Protein; 455 AA.
XX
AC   AAW33698;
XX
DT   30-APR-1998  (first entry)
XX
DE   AL-2-long (AL-2l) protein.
XX
OS   Homo sapiens.
XX
KW   AL-2l; AL-2; AL-2-long; human; treatment; neurological disorder; tumour;
KW   rheumatoid arthritis; wound healing; paralysis; angiogenesis; leukaemia;
KW   psoriasis; Alzheimer's disease; epilepsy.
XX
FH   Key            Location/Qualifiers
FH
FT   Peptide        1..26
FT                  /note= "signal peptide"
FT   Protein        27..455
FT                  /note= "mature protein"
FT   Domain         27..219
FT                  /note= "extracellular domain"
FT   Domain         220..245
FT                  /note= "hydrophobic transmembrane domain"
XX
PN   WO9740153-A1.
XX
```

Left column:

```
PD   30-OCT-1997.
XX
PF   17-APR-1997; 97WO-US06345.
XX
PR   19-APR-1996; 96US-0635130.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Caras IW;
XX
DR   WPI; 1997-535837/49.
DR   N-PSDB; AAV06354.
XX
PS   Claim 20; Fig 1A-C; 86pp; English.
XX
PT   Human AL-2 neurotrophic factor and related DNA - used to develop
PT   products for, e.g. treating neurologic disorders, angiogenesis
PT   disorders, tumours or rheumatoid arthritis or for wound healing
XX
CC   This is a AL-2-long (AL-2l) protein. The AL-2 is a novel Eph-related
CC   tyrosine kinase receptor ligand. AL-2 can be administered to patients in
CC   whom the nervous system has been damaged by trauma, surgery, stroke,
CC   ischaemia, infection, metabolic disease, nutritional deficiency,
CC   malignancy, or toxic agents, to promote the survival or growth of
CC   neurons. They can be used to treat motoneuron disorders such as
CC   amyotrophic lateral sclerosis (Lou Gehrig's disease), Bell's palsy, and
CC   various conditions involving spinal muscular atrophy, or paralysis. AL-2
CC   can be used to treat human neurodegenerative disorders, such as
CC   Alzheimer's disease, Parkinson's disease, epilepsy, demyelinating
CC   diseases such as multiple sclerosis, Huntingtons chorea, Down's syndrome,
CC   nerve deafness, Menier's disease, and other disorders of the cerebellum.
CC   AL-2 can be used as cognitive enhancer, to enhance learning particularly
CC   in dementias or trauma, since they can promote axonal outgrowth and
CC   synaptic plasticity, particularly of hippocampal neurons that express
CC   AL-2 binding Eph-family receptors and cortical neurons that express
CC   AL-2. AL-2 can also be used for wound healing, i.e. accelerating
CC   neovascularisation of, e.g. burns and ulcers. The encoding nucleic acids
CC   are useful in preparing antibodies that specifically bind to the AL-2
CC   protein. The antibodies and the AL-2 antagonists are useful in diagnosing
CC   and treating various neuronal disorders. AL-2 antagonists can be used
CC   for modulating angiogenesis. They can also be used for the treatment of
CC   tumours, acute myeloid leukaemia (AML), chronic myeloid leukaemia (CML),
CC   myelodysplastic syndrome (MDS), diabetic retinopathy, neovascular
CC   glaucoma, psoriasis and rheumatoid arthritis.
XX
SQ   Sequence 455 AA;

     Query Match       100.0%;  Score 2450;  DB 18;  Length 455;
     Best Local Similarity  100.0%;
     Matches  455;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy         1 MGPPHSGPGGVRVGALLLLGVLLGVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDL 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1 MGPPHSGPGGVRVGALLLLGVLLGVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDL 60

Qy        61 LCPRARPPGPHSSPNYEYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDLDLRFTIKFQEY 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 LCPRARPPGPHSSPNYEYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDLDLRFTIKFQEY 120

Qy       121 SPNIWGHEFRSHDYIIATSDGTREGLESLQGGVCLITRGMKVLLRVGQSPRGGAVPRKP 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 SPNIWGHEFRSHDYIIATSDGTREGLESLQGGVCLITRGMKVLLRVGQSPRGGAVPRKP 180

Qy       181 VSEMPMERDRGAAHSLEPGKENLPGDPTPSNATSRGAEGPLPPPSMPAVAGAAGGLALLLL 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 VSEMPMERDRGAAHSLEPGKENLPGDPTPSNATSRGAEGPLPPPSMPAVAGAAGGLALLLL 240

Qy       241 GVAGAGGANCWRRRRAKPSESRHPGPGSFGRGGSLGLGGGGWGPREAEPGELGIALRGG 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 GVAGAGGANCWRRRRAKPSESRHPGPGSFGRGGSLGLGGGGWGPREAEPGELGIALRGG 300

Qy       301 GAADPPFCPHYEKVSGDYGHPVYIVQDGPPQSPPNIYYTSISVLEWPILHTIQLFFMRSK 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       301 GAADPPFCPHYEKVSGDYGHPVYIVQDGPPQSPPNIYYTSISVLEWPILHTIQLFFMRSK 360

Qy       361 CSRVTTFLFPVQVITTSTCRMTSFSFTTLNPSMQACRAQMGEFRIRWCFWGDRILGTALF 420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 CSRVTTFLFPVQVITTSTCRMTSFSFTTLNPSMQACRAQMGEFRIRWCFWGDRILGTALF 420

Qy       421 VLVILILLGRLNMHQTTLLRQRASVEAEAGQHGPL 455
             |||||||||||||||||||||||||||||||||||
Db       421 VLVILILLGRLNMHQTTLLRQRASVEAEAGQHGPL 455

RESULT 2

ID   AAW33699 standard; Protein; 340 AA.
AAW33699
AC   AAW33699;
XX
DT   30-APR-1998  (first entry)
XX
DE   AL-2-short (AL-2s) protein.
XX
KW   AL-2s; AL-2; AL-2-short; human; treatment; neurological disorder; tumour;
KW   rheumatoid arthritis; wound healing; paralysis; angiogenesis; leukaemia;
KW   psoriasis; Alzheimer's disease; epilepsy.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..26
FT                   /note= "signal peptide"
FT   Protein         27..340
FT                   /note= "mature protein"
FT   Domain          27..219
FT                   /note= "extracellular domain"
FT   Domain          220..245
FT                   /note= "hydrophobic transmembrane domain"
XX
PN   WO9740153-A1.
XX
PD   30-OCT-1997.
XX
PF   17-APR-1997; 97WO-US06345.
XX
PR   19-APR-1996; 96US-0635130.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Caras IW;
XX
DR   WPI; 1997-535837/49.
DR   N-PSDB; AAV06355.
XX
PS   Claim 20; Fig 2A-B; 86pp; English.
XX
PT   Human AL-2 neurotrophic factor and related DNA - used to develop
PT   products for, e.g. treating neurologic disorders, angiogenesis
PT   disorders, tumours or rheumatoid arthritis or for wound healing
XX
CC   This is a AL-2-short (AL-2s) protein. The AL-2 is a novel Eph-related
CC   tyrosine kinase receptor ligand. AL-2 can be administered to patients in
CC   whom the nervous system has been damaged by trauma, surgery, stroke,
CC   ischaemia, infection, metabolic disease, nutritional deficiency,
CC   malignancy, or toxic agents, to promote the survival or growth of
CC   neurons. They can be used to treat motoneuron disorders such as
CC   amyotrophic lateral sclerosis (Lou Gehrig's disease), Bell's palsy, and
CC   various conditions involving spinal muscular atrophy, or paralysis. AL-2
CC   can be used to treat human neurodegenerative disorders, such as
CC   Alzheimer's disease, Parkinson's disease, epilepsy, demyelinating
CC   diseases such as multiple sclerosis, Huntingtons chorea, Down's syndrome,
CC   nerve deafness, Menier's disease, and other disorders of the cerebellum.
CC   AL-2 can be used as cognitive enhancer, to enhance learning particularly
CC   in dementias or trauma, since they can promote axonal outgrowth and
CC   synaptic plasticity, particularly of hippocampal neurons that express
```

```
CC   AL-2 binding Eph-family receptors and cortical neurons that express
CC   AL-2. AL-2 can also be used for wound healing, i.e. accelerating
CC   neovascularisation of, e.g. burns and ulcers. The encoding nucleic acids
CC   are useful in preparing antibodies that specifically bind to the AL-2
CC   protein. The antibodies and the AL-2 antagonists are useful in diagnosing
CC   and treating various neuronal disorders. They can also be used
CC   for modulating angiogenesis. AL-2 antagonists can be used
CC   tumours, acute myeloid leukaemia (AML), chronic myeloid leukaemia (CML),
CC   myelodysplastic syndrome (MDS), diabetic retinopathy, neovascular
CC   glaucoma, psoriasis and rheumatoid arthritis.
XX
SQ   Sequence   340 AA;

Qy   Query Match   75.1%;  Score 1841;  DB 18;  Length 340;
     Best Local Similarity   100.0%;  Pred. No. 4.4e-147;
     Matches  338;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy          1 MGPPHSGPGGVRVGALLLLGVLGLVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGPPHSGPGGVRVGALLLLGVLGLVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDL 60

Qy         61 LCPRARPPGPHSSPNYEFYKLYLVGGAQGRRCEAPPAPNLLTLCDRPDLDLRFTIKFQEY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LCPRARPPGPHSSPNYEFYKLYLVGGAQGRRCEAPPAPNLLTLCDRPDLDLRFTIKFQEY 120

Qy        121 SPNLWGHEFRSHHDYYIIATSDGTREGLESLCQGVCLTRGMKVLLRVGQSPRGGAVPRKP 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SPNLWGHEFRSHHDYYIIATSDGTREGLESLCQGVCLTRGMKVLLRVGQSPRGGAVPRKP 180

Qy        181 VSEMPMERDRGAAHSLEPGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAGGLALLLL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 VSEMPMERDRGAAHSLEPGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAGGLALLLL 240

Qy        241 GVAGAGGAMCWRRRRAKPSESRHPGPGSFGRGGSLGLGGGGNGPREAEPGELGIALRGG 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GVAGAGGAMCWRRRRAKPSESRHPGPGSFGRGGSLGLGGGGNGPREAEPGELGIALRGG 300

Qy        301 GAADPPFCPHYEKVSGDYGHPVVIVQDGPPQSPPNIYY 338
              ||||||||||||||||||||||||||||||||||||||
Db        301 GAADPPFCPHYEKVSGDYGHPVVIVQDGPPQSPPNIYY 338
```

```
RESULT 3
AAW31544
ID   AAW31544 standard; Protein; 340 AA.
XX
AC   AAW31544;
XX
DT   14-APR-1998 (first entry)
XX
DE   Human cytokine Lerk-8.
XX
KW   Lerk-8; cytokine; human; hek; elk; receptor tyrosine kinase;
KW   ligand; neurodegenerative disease; wound healing;
KW   neovascularisation; diagnosis; therapy.
XX
OS   Homo sapiens.
XX
XX
FH   Key             Location/Qualifiers
FH
FT   Peptide         1..27
FT                   /label= Sig_peptide
FT   Protein         28..340
FT                   /label= Mat_protein
FT   Domain          28..224
FT                   /note= "extracellular domain"
FT   Domain          225..251
FT                   /note= "transmembrane domain"
FT   Domain          252..340
FT                   /note= "cytoplasmic domain"
FT   Modified-site   210..212
FT                   /note= "N-glycosylated"
FT   Misc-difference 325
FT                   /note= "residue 325 is Leu in Lerk-8 variant"
```

```
XX
PN   WO9736919-A2.
XX
PD   09-OCT-1997.
XX
PF   19-MAR-1997;    97WO-US04533.
XX
PR   21-MAR-1996;    96US-0621146.
XX
PA   (IMMV ) IMMUNEX CORP.
XX
PI   Cerretti DP;
XX
DR   WPI; 1997-503043/46.
DR   N-PSDB; AAT89519.
XX
PT   New isolated cytokine, Lerk-8 - binds to the hek and elk receptor
PT   tyrosine kinases, used to develop products for diagnosis and therapy
XX
PS   Claim 3; Page 32-33; 37pp; English.
XX
CC   This protein sequence comprises a novel human cytokine designated
CC   Lerk-8. The amino acid sequence was deduced from a human foetal
CC   brain cDNA clone (see AAT89519). Lerk-8 binds to the cell surface
CC   receptors hek and elk, which are members of the eph/elk family of
CC   receptor tyrosine kinases. Lerk-8 polypeptides, especially soluble
CC   polypeptides comprising amino acid residues -27 to 142-197 of the
CC   full-length protein, can be expressed in transformed host cells.
CC   These polypeptides can be used to purify hek or elk proteins, and
CC   to purify or identify cells that express hek or elk on the surface.
CC   Such cells can be used in various in vitro studies or in vivo
CC   procedures, e.g. neural cells expressing elk can be administered to
CC   a mammal afflicted with a neurodegenerative disorder. The Lerk-8
CC   polypeptides can also be used to deliver diagnostic or therapeutic
CC   agents to these cells (e.g. leukaemia cells). The Lerk-8 DNA and
CC   polypeptides can be used to: treat disorders mediated by
CC   defective or insufficient amounts of Lerk-8; to treat disorders
CC   such as injury to neural tissue or neurologic disease; to promote
CC   angiogenesis; and for wound healing or stimulating
CC   neovascularisation of grafted tissues.
XX
SQ   Sequence   340 AA;

Qy   Query Match   75.1%;  Score 1841;  DB 18;  Length 340;
     Best Local Similarity   100.0%;  Pred. No. 4.4e-147;
     Matches  338;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy          1 MGPPHSGPGGVRVGALLLLGVLGLVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGPPHSGPGGVRVGALLLLGVLGLVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDL 60

Qy         61 LCPRARPPGPHSSPNYEFYKLYLVGGAQGRRCEAPPAPNLLTLCDRPDLDLRFTIKFQEY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LCPRARPPGPHSSPNYEFYKLYLVGGAQGRRCEAPPAPNLLTLCDRPDLDLRFTIKFQEY 120

Qy        121 SPNLWGHEFRSHHDYYIIATSDGTREGLESLCQGVCLTRGMKVLLRVGQSPRGGAVPRKP 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SPNLWGHEFRSHHDYYIIATSDGTREGLESLCQGVCLTRGMKVLLRVGQSPRGGAVPRKP 180

Qy        181 VSEMPMERDRGAAHSLEPGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAGGLALLLL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 VSEMPMERDRGAAHSLEPGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAGGLALLLL 240

Qy        241 GVAGAGGAMCWRRRRAKPSESRHPGPGSFGRGGSLGLGGGGNGPREAEPGELGIALRGG 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GVAGAGGAMCWRRRRAKPSESRHPGPGSFGRGGSLGLGGGGNGPREAEPGELGIALRGG 300

Qy        301 GAADPPFCPHYEKVSGDYGHPVVIVQDGPPQSPPNIYY 338
              ||||||||||||||||||||||||||||||||||||||
Db        301 GAADPPFCPHYEKVSGDYGHPVVIVQDGPPQSPPNIYY 338
```

RESULT 4

```
ID   AAW10637 standard; Protein; 340 AA.
XX
AC   AAW10637;
XX
DT   23-JUN-1997 (first entry)
XX
DE   NLERK2 ligand for eph-related kinase.
XX
KW   LERK; ligand for eph-related kinase; ERK; NLERK2;
KW   receptor protein tyrosine kinase; cell proliferation;
KW   cell differentiation; cell survival; nerve cell.
XX
OS   Homo sapiens.
XX
FH   Key            Location/Qualifiers
FT   Peptide        1..29
FT                  /label= Sig_peptide
FT   Protein        30..340
FT                  /label= Mat_protein
FT   Domain         227..251
FT                  /label= Transmembrane_domain
FT   Modified-site  210
FT                  /label= N-glycosylation_site
XX
PN   WO9704091-A1.
XX
PD   06-FEB-1997.
XX
PF   19-JUL-1996; 96WO-AU00460.
XX
PR   05-FEB-1996; 96AU-0007890.
PR   20-JUL-1995; 95AU-0004263.
PR   27-NOV-1995; 95AU-0006847.
PR   22-DEC-1995; 95AU-0007299.
XX
PA   (AMRA-) AMRAD OPERATIONS PTY LTD.
XX
PI   Nicola NA;
XX
DR   WPI; 1997-132632/12.
DR   N-PSDB; AAT60966.
XX
PT   Nucleic acid mol. encoding ligand for eph-related kinase - useful
PT   for treatment of, pref. neuronal, cells to increase survival,
PT   proliferation and differentiation.
XX
PS   Claim 16; Page 37-39; 71pp; English.
XX
CC   A novel human ligand for eph-related kinase (LERK) is designated
CC   NLERK2 (AAW10637). It is encoded by a cDNA clone (AAT60966) obtd.
CC   from a human foetal brain cDNA library. The novel receptor ligand
CC   can be expressed in transformed host cells and used in methods
CC   for regulating the development, maintenance or regeneration of
CC   different cells (e.g. neurons) and tissues in vivo and in vitro.
CC   Soluble NLERK2 peptides can be used to treat injury, disease or
CC   abnormality in the nervous system, and membrane-bound NLERK2 to
CC   modulate proliferation, different or survival e.g. in grafting
CC   procedures or transplantation. NLERK2 can also be used to raise
CC   antibodies for use in immunotherapy, and to detect anti-NLERK2
CC   antibodies that may occur in some autoimmune diseases.
XX
SQ   Sequence 340 AA;

  Query Match             75.1%;  Score 1841;  DB 18;  Length 340;
  Best Local Similarity   100.0%;
  Matches  338;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGPPHSGPGGVRVGALLLLGVTLGLVSGLSLEPVYWNSANKRFQAEGGYYLYPQIGDRLDL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGPPHSGPGGVRVGALLLLGVTLGLVSGLSLEPVYWNSANKRFQAEGGYYLYPQIGDRLDL 60

Qy         61 LCPRARPPGPHSSPNYEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDLDLRFTIKFQEY 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LCPRARPPGPHSSPNYEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDLDLRFTIKFQEY 120

Qy        121 SPNLWGHEFRSHHDYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVCQSPRGGAVPRKP 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SPNLWGHEFRSHHDYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVCQSPRGGAVPRKP 180

Qy        181 VSEMPMERDRGAAHSLEPGKENLPGDPTSNATSRGAEGPLPPSMPAVAGAAGGLALLLL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 VSEMPMERDRGAAHSLEPGKENLPGDPTSNATSRGAEGPLPPSMPAVAGAAGGLALLLL 240

Qy        241 GVAGAGGAMCWRRRAKPSESRHPGPGSFGRGGSLGLGGGGMGPREAEPGELGIALRGG 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GVAGAGGAMCWRRRAKPSESRHPGPGSFGRGGSLGLGGGGMGPREAEPGELGIALRGG 300

Qy        301 GAADPPFCPHYEKVSGDYGHPVYIVQDGPPQSPFNIYY 338
              ||||||||||||||||||||||||||||||||||||||
Db        301 GAADPPFCPHYEKVSGDYGHPVYIVQDGPPQSPFNIYY 338


RESULT 5
AAW17081
ID   AAW17081 standard; Protein; 340 AA.
XX
AC   AAW17081;
XX
DT   09-AUG-1997 (first entry)
XX
DE   EPH family ligand Efl-6.
XX
KW   Efl-6; Eph; Elk; receptor tyrosine kinase; signal transduction;
KW   ligand; neurological disease.
XX
OS   Homo sapiens.
XX
FH   Key            Location/Qualifiers
FT   Peptide        1..24
FT                  /label= Sig_peptide
FT   Protein        25..340
FT                  /label= Mat_protein
FT   Domain         225..249
FT                  /label= Transmembrane_domain
FT   Misc-difference 166
FT                  /label= Gln, Arg
XX
PN   WO9715667-A1.
XX
PD   01-MAY-1997.
XX
PF   25-OCT-1996; 96WO-US17201.
XX
PR   25-OCT-1995; 95US-0007015.
XX
PA   (REGE-) REGENERON PHARM INC.
XX
PI   Davis S, Gale NW, Yancopoulos GD;
XX
DR   WPI; 1997-259021/23.
DR   N-PSDB; AAT69808.
XX
PT   New nucleic acid encoding Efl-6 ligand protein - used for promoting
PT   growth and proliferation of neuronal cells and in drug screening
XX
PS   Claim 2; Fig 1; 36pp; English.
XX
CC   A novel ligand (AAW17081), designated Efl-6 (or Eph transmembrane
CC   tyrosine kinase family ligand 6), binds to the Elk, Nuk/Cek5,
CC   Hek2/Sek4, Htk and Seki receptors on cells. Its amino acid
CC   sequence was deduced from a human frontal cortex cDNA clone
CC   (AAT69808). Recombinant Elf-6, truncated soluble polypeptides
CC   comprising the extracellular domain of Elf-6, and Elf-6
CC   ligandbodies comprising soluble Efl-6 and the Fc portion of IgG can
CC   be expressed in host cells. These can be used to support neuronal
```

CC   and other Eph receptor-bearing cell populations for treatment of
CC   neurological disorders, in drug screening and to raise diagnostic
CC   antibodies.
XX
SQ   Sequence 340 AA;

Query Match   74.9%; Conservative 0; Pred. No. 1.4e-146;
Best Local Similarity   99.7%;
Matches 337; Score 1835; DB 18; Length 340;
                Mismatches 1; Indels 0; Gaps 0;

Qy   1 MGPPHSGPGGVRVGALLILGVLGLVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDL   60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MGPPHSGPGGVRVGALLILGVLGLVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDL   60

Qy   61 LCPRARPPGPHSSSPNYEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDLDLRFTIKFQEY 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LCPRARPPGPHSSSPNYEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDLDLRFTIKFQEY 120

Qy 121 SPNLWGHEFRSHHDYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKP 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 SPNLWGHEFRSHHDYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKP 180

Qy 181 VSEMPMERDRGAAHSLEPGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAGGLALLLL 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 VSEMPMERDRGAAHSLEPGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAGGLALLLL 240

Qy 241 GVAGAGGAMCWRRRRAKPSESRHPGPGSFGRGGSLGLGGGGGMGPREAEPGELGIALRGG 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 GVAGAGGAMCWRRRRAKPSESRHPGPGSFGRGGSLGLGGGGGMGPREAEPGELGIALRGG 300

Qy 301 GAADPPFCPHYEKVSGDYGHPVVIVQDGPPQSPPNIIY 338
         ||||||||||||||||||||||||||||||||||||||
Db 301 GAADPPFCPHYEKVSGDYGHPVVIVQDGPPQSPPNIIY 338

RESULT 6
AAW46615
ID   AAW46615   standard; Protein; 340 AA.
XX
AC   AAW46615;
XX
DT   06-JUL-1998 (first entry)
XX
DE   Human transmembrane ligand Elk-L3.
XX
KW   Elk-L3; Elk-related receptor tyrosine kinase; transmembrane ligand;
KW   human; signal transduction; axonogenesis; nerve cell; neurone;
KW   Alzheimer's disease; Parkinson's disease; Huntingdon's disease;
KW   demyelination; multiple sclerosis; amyotrophic lateral sclerosis;
KW   nervous system infection; Wernicke's disease; trauma; ischaemia;
KW   stroke; nutritional polyneuropathy; progressive supranuclear palsy;
KW   Shy Drager's syndrome; multistem degeneration;
KW   olivo ponto cerebellar atrophy, peripheral nerve damage.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   Misc-difference 166
FT   /label= Gln, Arg
FT   Domain   225..249
FT   /note= "transmembrane domain"
XX
PN   WO9801548-A1.
XX
PD   15-JAN-1998.
XX
PF   04-JUL-1997; 97WO-CA00473.
XX
PR   05-JUL-1996; 96US-0021272.
XX
PA   (MOUN) MOUNT SINAI HOSPITAL CORP.
XX
PI   Holland S, Mbamalu G, Pawson T;

XX
DR   WPI; 1998-101047/09.
DR   N-PSDB; AAV16097.
XX
PT   Modulating transmembrane ligand for an Elk-related receptor tyrosine
PT   kinase - by formation of a complex between an oligomerised
PT   Elk-related receptor tyrosine kinase and a transmembrane ligand
XX
PS   Disclosure; Fig 5A; 40pp; English.
XX
CC   This polypeptide comprises human Elk-L3, a transmembrane ligand of
CC   Elk-related receptor tyrosine kinase (ERRTK). A novel method of
CC   modulating the biological activity of, or for affecting a pathway
CC   regulated by, a transmembrane ligand for an ERRTK in a cell
CC   expressing the transmembrane ligand comprises forming a complex
CC   between a purified and isolated oligomerised ERRTK, or an isoform
CC   or an extracellular domain of the ERRTK, and the transmembrane
CC   ligand expressed on the cell. The complex can also be used for
CC   evaluating a substance for its ability to modulate the biological
CC   activity of a transmembrane ligand for an ERRTK, and to identify
CC   substances that affect or modulate a pathway regulated by ERRTK.
CC   A purified and isolated oligomerised ERRTK can be used in the
CC   preparation of a medicament for modulating neuronal development or
CC   regeneration in a subject, or in a medicament for modulating
CC   axonogenesis in a subject (all claimed). The substances identified
CC   by the methods can be used to modulate axonogenesis, nerve cell
CC   interactions and regeneration, to treat diseases and conditions
CC   involving trauma and injury to the nervous system, such as
CC   Alzheimer's disease, Parkinson's disease, Huntingdon's disease,
CC   lateral sclerosis, bacterial and viral infections of the nervous
CC   system, deficiency diseases, such as Wernicke's disease and
CC   demyelinating diseases, such as multiple sclerosis, amyotrophic
CC   nutritional polyneuropathy, progressive supranuclear palsy,
CC   Shy Drager's syndrome, multistem degeneration and olivo ponto
CC   cerebellar atrophy, peripheral nerve damage, trauma, and ischaemia
CC   resulting from stroke.
XX
SQ   Sequence 340 AA;

Query Match   74.6%; Conservative 0; Pred. No. 6.7e-146;
Best Local Similarity   99.4%;
Matches 336; Score 1827; DB 19; Length 340;
                Mismatches 2; Indels 0; Gaps 0;

Qy   1 MGPPHSGPGGVRVGALLILGVLGLVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDL   60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MGPPHSGPGGVRVGALLILGVLGLVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDL   60

Qy   61 LCPRARPPGPHSSSPNYEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDLDLRFTIKFQEY 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LCPRARPPGPHSSSPNYEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDLDLRFTIKFQEY 120

Qy 121 SPNLWGHEFRSHHDYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKP 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 SPNLWGHEFRSHHDYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKP 180

Qy 181 VSEMPMERDRGAAHSLEPGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAGGLALLLL 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 VSEMPMERDRGAAHSLEPGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAGGLALLLL 240

Qy 241 GVAGAGGAMCWRRRRAKPSESRHPGPGSFGRGGSLGLGGGGGMGPREAEPGELGIALRGG 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 GVAGAGGAMCWRRRRAKPSESRHPGPGSFGRGGSLGLGGGGGMGPREAEPGELGIALRGG 300

Qy 301 GAADPPFCPHYEKVSGDYGHPVVIVQDGPPQSPPNIIY 338
         ||||||||||||||||||||||||||||||||||||||
Db 301 GAADPPFCPHYEKVSGDYGHPVVIVQDGPPQSPPNIIY 338

RESULT 7
AAW00287
ID   AAW00287   standard; Protein; 334 AA.
XX
AC   AAW00287;

XX
DT   19-JAN-1997 (first entry)
XX
DE   Mouse Eph receptor ligand ELF-2.
XX
KW   Eph receptor; ligand; ELF-2; tyrosine kinase; signal transduction;
KW   organogenesis; oncogenesis; tumour; neurological disorder;
KW   diagnosis; gene therapy.
XX
OS   Mus sp.
XX
FH   Key   Location/Qualifiers
FH
FT   Domain   1..224
FT   /label= Extracellular_domain
FT   Binding-site   31..155
FT   /label= Receptor_binding_site
FT   Domain   226..251
FT   /label= Transmembrane_domain
XX
PN   WO9626958-A2.
XX
PD   06-SEP-1996.
XX
PF   23-FEB-1996; 96WO-US02673.
XX
PR   27-FEB-1995; 95US-0355415.
XX
PA   (HARD) HARVARD COLLEGE.
XX
PI   Bergemann AD, Flanagan JG;
XX
DR   WPI; 1996-433391/43.
DR   N-PSDB; AAT40230.
XX
PT   Eph receptor ligand, ELF-2, and DNA encoding it - used to treat or
PT   prevent neurological diseases, and to modulate binding of ELF-2 to
PT   Eph receptor, e.g. to prevent or treat tumour formation.
XX
PS   Claim 6; Fig 1A-B; 50pp; English.
XX
CC   Mouse Eph receptor ligand ELF-2 (AAW00287) is strongly expressed in
CC   the anterior hindbrain and newly-forming somites of embryos at the
CC   early organogenesis stage of development. It is important in
CC   cellular communication during pattern formation. Its amino acid
CC   sequence was deduced from a cDNA clone (AAT40230) isolated from a
CC   newborn mouse brain cDNA library. The ELF-2 ligand can be used
CC   to alter neurological development, oncogenesis and growth
CC   regulation, to modulate binding of ELF-2 to the Eph receptor, and
CC   in diagnostic assays.
XX
SQ   Sequence   334 AA;

Query Match   25.7%; Score 628.5; DB 17; Length 334;
Best Local Similarity   41.8%; Pred. No. 9.5e-45;
Matches 137; Conservative   49; Mismatches 129; Indels 13; Gaps 5;

Qy   14 GALILLGVLGLVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDLLCPRARPPGHSS 73
              |:|.|||||||||||:||||:|:|||||||||||||||||||||||||||||:
Db   15 GLIMVLCRTAISRSIVLEPIYWNSSNSKFLPGQGLVLYPQIGDKLDIICPKV---DSKTV 71

Qy   74 PNYEFYKLIVLVGGAQGRRCEAPAPANILLITCDRPDLDLRFTIKPQEYSPNLWGHEFRSHH 133
              |||:|:|||.|:|||
Db   72 GQYEYKVXVWVDKDQADRCTIKKENTPLLNCARPDQDVKFTIKPQEFSPNLWGLEFQKNK 131

Qy   134 DYYIIATSDGTREGLESLQGGVCLITRGMKVLLRVGQ---SPRGGAVPRKPVSEMPMER-DR 190
              ||:|||:||||:|:||...|:||:||..:||...
Db   132 DYYIISTSNGSLEGLDNQEGGVCQTRAMKILMKVGQDASSGSARNHGPTRRPELEAGTN 191

Qy   191 GAAHSLEPGKENLPGDPPSNATSRGAEGPLPPPSMPAVAGAAGGLALLLLGVAGAGGAMC 250
              |||.||:|||:|..:|||:...|
Db   192 GRSSTTSPFVKPNPGSSTDGNSAGHSGNNLLGSEVALFAGIASGCIFIVIILTLVVLLL 251

Qy   251 WRRRAKPSESRHPGPGSFGRGGSLGLGGGGMGPREAPEGELGIALRGGGADPPFGCH 310

Db   252 KYRRRRKHSPQHTTLSLSTLATPKRGGNN-----NGSEPSDVIIPLR---TADSVFCPH 304
              |||.:|.| |||.:|||..|.|...|| :||:|.|||.| ||.||||

Qy   311 YEKVSGDYGHPVYIVQDGPQGSPENIYY 338
              |||:||||||||||:..||.|||.|||
Db   305 YEKVSGDYGHPVYIVQEMPPQSPANIYY 332

RESULT 8
ID   AAR92742 standard; Protein; 336 AA.
XX
AC   AAR92742;
XX
DT   21-MAY-1996 (first entry)
XX
DE   Murine hepatoma transmembrane kinase receptor ligand.
XX
KW   Hepatoma transmembrane kinase; Htk; receptor; ligand;
KW   tyrosine kinase; neurodegenerative disease.
XX
OS   Mus musculus.
XX
PN   WO9602645-A2.
XX
PD   01-FEB-1996.
XX
PF   14-JUL-1995; 95WO-US08812.
XX
PR   20-JUL-1994; 94US-0277722.
XX
PA   (GETH) GENENTECH INC.
XX
PI   Bennett BD, Matthews W;
XX
DR   WPI; 1996-105907/11.
DR   N-PSDB; AAT16470.
XX
PT   Ligand for the hepatoma trans-membrane kinase receptor - useful for
PT   stimulating and inhibiting cells carrying the receptor, e.g. for
PT   treating neuro-degenerative disease
XX
PS   Claim 5; Fig 1(A-D); 88pp; English.
XX
CC   Mouse (AAT16470) and human (AAT16471) Htk ligand which bind to, and
CC   activate, the Htk receptor, have been identified in a variety of
CC   tissues using a soluble Htk-Fc fusion protein.
CC   The predicted mol.wt. of the murine Htk ligand protein following a
CC   signal peptide cleavage is 34 kD with an estimated pI of 8.9.
CC   The murine and human ligands show 96% homology at the amino acid
CC   level.
CC   The DNA is used to produce recombinant ligands; for tissue-
CC   specific typing (partic. as a marker for breast cancer) and as a
CC   marker for human chromosome 13. The ligands (partic. in soluble
CC   form) are used to activate the tyrosine kinase domain of the
CC   Htk receptor, i.e. to stimulate or inhibit growth, differentiation,
CC   and/or activation of cells contg. the receptor, e.g. treatment
CC   of neurodegenerative diseases, since they are strongly expressed
CC   in the cerebral cortex, hippocampus, striatum and cerebellum.
CC   The ligands are also useful as a control or standard in assays,
CC   for generation of antibodies, as a mol. wt. marker, for growth
CC   in vitro of Htk-receptor positive cells, as research agent,
CC   in screening, etc.
XX
SQ   Sequence   336 AA;

Query Match   25.7%; Score 628.5; DB 17; Length 336;
Best Local Similarity   41.8%; Pred. No. 9.6e-45;
Matches 137; Conservative   49; Mismatches 129; Indels 13; Gaps 5;

Qy   14 GALILLGVLGLVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDLLCPRARPPGHSS 73
              |:|.|||||||||||:||||:|:|||||||||||||||||||||||||||||:
Db   17 GLIMVLCRTAISRSIVLEPIYWNSSNSKFLPGQGLVLYPQIGDKLDIICPKV---DSKTV 73

Qy   74 PNYEFYKLVIVGGAQGRRCEAPPAPNLLLTCDRPDLDRFTIKPQEYSPNLWGHEFRSHH 133
              |:||:||:||||:||||||||:||||::||||||||:||:|||||:|:
Db   74 GQYEYYKVYMVDKQDADRCTIKKENTPLLNCARPDQODVKFTIKPQEBFSPNLWLEFQKNK 63

Qy   134 DYYIIATSDGTREGLESLQGGVCLLTRGMKVLLRVGQ--SPRGGAVPRKPVSEMPMER-DR 190
              |:|||:|||||||:||:|||||||||||||||:||| ::||:::||||||| ||
Db   134 DYYIISTNSGSLEGLDNQEGGVCQTRAMKILMKVGQDASSGSARNHGPTRRPELEAGTN   133

Qy   191 GAAHSLEPGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAGGLALLLLGVAGAGGAMC 250
              |:|||||||::||||||||:||||||||||||||||||:|:||||||||||||||:
Db   194 GRSSTTSPFVKPNFGGSSTDGNSAGHSGNNLLGSEVALFAGIASGCCIFIVIIIILVVLLL 253

Qy   251 WRRRRAKPSESRHPGPGSFQRGGSLGLGGGGGMGPREAEPGELGIALRGGGAADPPPFCPH 310
              |||||||||||||:|||:|||||:|||||::::::::::||:||||||||||||||
Db   254 KYRRRHRKHSPQHTTTLSLTLATPKRGGNN----NGSEPSDVIIPLR---TADSVFCPH 306

Qy   311 YEKVSGDYGHPVYIVQDGPPQSPENIYY 338
              |||||||||||||||||||||||||||
Db   307 YEKVSGDYGHPVIIVQEMPPQSPANIYY 334

RESULT 9
AAR55059
XX
ID   AAR55059 standard; Protein; 346 AA.
XX
AC   AAR55059;
XX
DT   28-JAN-1995 (first entry)
XX
DE   Elk tyrosine kinase receptor ligand.
XX
KW   Vectors; elk-L protein; elk; ligands; cell growth; differentiation.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FH
FT   Peptide   1..24
FT   /note= "signal peptide"
FT   Protein   25..346
FT   /note= "mature elk-L protein"
XX
PN   WO9411384-A.
XX
PD   26-MAY-1994.
XX
PF   15-NOV-1993; 93WO-US10955.
XX
PR   13-NOV-1992; 92US-0977693.
XX
PA   (IMMV) IMMUNEX CORP.
XX
PI   Baum PR, Beckmann MP, Lyman S;
XX
DR   WPI; 1994-183415/22.
DR   N-PSDB; AAQ65486.
XX
PT   New DNA encoding ligand for elk tyrosine kinase receptor - also
PT   related polypeptides, vectors, antibodies and probes, useful e.g.
PT   in studying cell differentiation or growth
XX
PR   Claim 7; Page 30; 35pp; English.
XX
CC   The sequence is that of the elk-L protein able to bind elk, a
CC   tyrosine kinase receptor. The DNA may be incorporated into vectors
CC   which can used to study the role of elk and its ligands in cell
CC   growth and differentiation.
XX
SQ   Sequence   346 AA;

Query Match   25.4%; Score 623; DB 15; Length 346;
Best Local Similarity   39.2%; Pred. No. 2.9e-44;
Matches 143; Conservative 48; Mismatches 116; Indels 58; Gaps 9;

Qy   8 PGGVRVGALLLLGVILGLVSGLL-----SLEPYTWNSANKRFQAEGGYVLLYPQIGDRLDLL 61
              |:||||||||||||||||||| ||||||||||||||||||:|:|||::|||:|||
Db   4 PGQRWLGRWLVAMVWWALCRLATPLAQNLEPVSWSSLNPKFLSGKGLIVIYPKIGDKLDII 63

Qy   62 CPRARPGPHSSPNYEFYKLVIVGGAQGRRCEAPPAPNLLLTCDRPDLDRFTIKPQEYS 121
              |||:||||:||:||||||||||||||||||||||||||||||||||||||||||
Db   64 CPPAEAGRP------YEYYKLVVRPEQQAACSTVLDNVILVTCNRPEQEIRFTIKPQEFS 118

Qy   122 PNLWGHEFRSHHDYYIIATSDGTREGLESLQGGVCLLTRGMKVLLRVGQSPRGGAVPRKPV 181
              |||||||||||||||||||:||||||||:||:|||||||||||||||||:|:||:||||
Db   119 PNTMGLEFKKHHDYYIISTNSGSLEGLENREGGVCRTRTMKIIMKVGQDPNAVTPEQLTT 178

Qy   182 SEMPMERDRGAAHSLE-PGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAGGLA---- 236
              |:||:|| ||::||| ||::||||||||:||||:|||||||||||||:|:|||||
Db   179 SRPSKEADNTVKMATQAPGSRGSLGDSDGKHFTVNQEEKSGP----GASGGSSGDPD 231

Qy   237 ------LLLLGVAGAGGA----------MCWRRRRAKPSESRHPGPGSFQRGGSLGL 277
                    |||||||||||| ::|||||||||||||:|||:|||||:|||
Db   232 GFFNSKVALFAAVGAGVCVIFLLIIIIFTVLLLIKLRLRKHRKHTQQ------RAAALSL 282

Qy   278 -----GGGGGMGPREAEPGELGIALRGGGAADPPPFCPHYEKVSGDYGHPVYIVQDGPPQSP 333
                   |||||::::::::::||:|||||||||||||||||||||||||||||||||||||
Db   283 STLASPKGGSGTRAGTEPSDIIIPLR----TTENNYCPHYEKVSGDYGHPVIIVQEMPPQSP 339

Qy   334 PNIYY 338
              |||||
Db   340 ANIYY 344

RESULT 10
AAR91930
XX
ID   AAR91930 standard; Protein; 346 AA.
XX
AC   AAR91930;
XX
DT   11-DEC-1996 (first entry)
XX
DE   Human cytokine elk-ligand (elk-L).
XX
KW   Human; cytokine; elk-ligand; elk-L; tyrosine kinase receptor;
KW   neurotrophic; neuroprotective; placenta; radiolabelled probe;
KW   treatment; neural tissue; excito-toxicity; injury; disorder;
KW   neural culture reagent; immunogenic fragment; antibody.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FH
FT   Peptide   1..24
FT   /label= sig_peptide
FT   Peptide   25..346
FT   /label= mat_peptide
XX
PN   US5512457-A.
XX
PD   30-APR-1996.
XX
PF   13-NOV-1992; 92US-0977693.
XX
PR   15-MAR-1994; 94US-0213403.
PR   13-NOV-1992; 92US-0977693.
XX
PA   (IMMV) IMMUNEX CORP.
XX
PI   Baum PR, Beckmann MP, Carpenter MK, Lyman S;
XX
DR   WPI; 1996-229866/23.
DR   N-PSDB; AAT28770.
XX
PT   DNA coding for neurotrophic human elk ligand cytokine - useful as
PT   probe to isolate other elk ligand sequences
XX
PS   Claim 1; Columns 27-30; 18pp; English.

XX
CC The present sequence is the human cytokine elk-ligand (elk-L),
CC which binds a member of the tyrosine kinase receptor family. Elk-L
CC exhibits neurotrophic and neuroprotective properties, and has a
CC calculated mol. wt. 35180 daltons and a pi of 9.006. The elk-L
CC cDNA, isolated from a human placental cDNA library, can be
CC radiolabelled and used as a probe for isolating other mammalian
CC elk-L cDNA. Elk-L can be used to treat neural tissue disorders,
CC patic. excito-toxicity associated injuries or disorders, and as a
CC neural culture reagent, while immunogenic fragments of elk-L can be
CC used to generate specific anti-elk-L antibodies.
XX
SQ Sequence   346 AA;

Query Match   25.4%; Score 623; DB 17; Length 346;
Best Local Similarity   39.2%;
Matches 143; Conservative 48; Pred. No. 2.9e-44; Mismatches 116; Indels 58; Gaps 9;

Qy   8 PGGVRVGALLLLGVLGLVSGL-------SLEPVYWNSANKRFQAEGGVYVLYPQIGDRLDLL 61
               |||| ||.||:. ||| || | | |:|::: |: .|:.|: ||||:: :||
Db   4 PGQRWLGKWLVAMVWWALCRLATPLAKNLEPVSWSSLNPKFLSGKGLVIYPKIGDKLDII 63

Qy   62 CPRARPPGPHSSSPNYEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDLDLRFTIKFQEYS 121
               ||| || |.|||:.| |.:|.|| |:|||| |.| | .||:| || |:|||||||:
Db   64 CPRAEAGRP-----YEYYKLYLVRPEQAAACSTVLDPNVLVTCNRPEQEIRFTIKFQEFS 118

Qy   122 PNLWGHEFRSHHDYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKPV 181
               ||||| ||.||||||||||||| || ||||:|||||||||||||:|||:|||||||||
Db   119 PNYMGLEFKKHHDYYITSTSNGSLEGLENRGGVCRTRTMKIIMKVGQDPNAVTPEQLTT 178

Qy   182 SEMPMERDRGAAHSLE-PGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAGGLA---- 236
               || |.|.|.. |:| |||:|||||::|.|:| |: | ||:|.| ||.|| |
Db   179 SRPSKEADNTVKWATQAPGSRGSLGDSDGKHETVNQEEKSGP--------GASGGSSGDPD 231

Qy   237 ------LILLGVAGAGGA------MCWRRRRAKPSESRHPGPGSFGRGGSLGL 277
                      |||:| | || ||:| | |: |:||| |||| |||
Db   232 GFFNSKVALFAAVGAGCVIFLLIIFLTVLLLKLRKRHRKHTQQ-------RAAALSL 282

Qy   278 --GGGGGMGPREAEPGELGIALRGGGAADPFCPHYEKVSGDYGHPVYIYVQDGPPQSP 333
               | : ||.:|||||| : |||||||::.||||||||||||||||.: ||| |||
Db   283 STLASPKGGSGTAGTEPSDIIIPLR---TTENNYCPHYEKVSGDYGHPVYIYVQEMPPQSP 339

Qy   334 PNIIY 338
               |||||
Db   340 ANIYY 344

RESULT 11
ID AAW36055 standard; Protein; 346 AA.
XX
AC AAW36055;
XX
DT 06-MAR-1998 (first entry)
XX
DE Human elk-L protein.
XX
OS Homo sapiens.
XX
KW extracellular domain; immunoglobulin; neurological disease.
KW Human; elk-L; cytokine; ligand; tyrosine kinase receptor; fusion protein;
XX
FH Key   Location/Qualifiers
FT Peptide   1..24
FT   /note= "signal peptide"
FT Protein   25..346
FT   /note= "mature protein"
FT Domain   25..237
FT   /note= "extracellular domain; this region is used to
FT   generate a fusion protein with the Fc portion
FT   of the human immunoglobulin G1"
FT Domain   238..265
FT   /note= "transmembrane domain"
FT Domain   266..346
FT   /note= "intracellular domain"
FT Modified-site   139..141
FT   /note= "Asn is N-glycosylated"
FT Cleavage-site   265..267
FT   /note= "KEX2 protease cleavage site"
FT Cleavage-site   267..268
FT   /note= "KEX2 protease cleavage site"
FT Cleavage-site   270..271
FT   /note= "KEX2 protease cleavage site"
XX
PN US5670625-A.
XX
PD 23-SEP-1997.
XX
PF 02-JUN-1995; 95US-0460741.
XX
PR 15-MAR-1994; 94US-0213403.
PR 13-NOV-1992; 92US-0977693.
PR 02-JUN-1995; 95US-0460741.
XX
PA (IMMV) IMMUNEX CORP.
XX
PI Baum PR, Beckmann MP, Lyman S;
XX
DR WPI; 1997-479524/44.
DR N-PSDB; AAI97976.
XX
PS Claim 1; Columns 27-30; 18pp; English.
XX
CC This is the amino acid sequence of the human elk-L protein, a new
CC cytokine that is the ligand for the elk tyrosine kinase receptor. The
CC extracellular domain of the protein (amino acids 1-213) is used to
CC generate a fusion protein comprising the Fc polypeptide of the human
CC immunoglobulin G1 (extending from the hinge region to the C-terminus).
CC The fusion protein (which has the same activities as the natural elk-L
CC protein) has neuroprotective and neurotrophic activity so is potentially
CC useful for treating a wide range of neurological diseases.
XX
SQ Sequence   346 AA;

Query Match   25.4%; Score 623; DB 18; Length 346;
Best Local Similarity   39.2%;
Matches 143; Conservative 48; Pred. No. 2.9e-44; Mismatches 116; Indels 58; Gaps 9;

Qy   8 PGGVRVGALLLLGVLGLVSGL-------SLEPVYWNSANKRFQAEGGVYVLYPQIGDRLDLL 61
               |||| ||.||:. ||| || | | |:|::: |: .|:.|: ||||:: :||
Db   4 PGQRWLGKWLVAMVWWALCRLATPLAKNLEPVSWSSLNPKFLSGKGLVIYPKIGDKLDII 63

Qy   62 CPRARPPGPHSSSPNYEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDLDLRFTIKFQEYS 121
               ||| || |.|||:.| |.:|.|| |:|||| |.| | .||:| || |:|||||||:
Db   64 CPRAEAGRP-----YEYYKLYLVRPEQAAACSTVLDPNVLVTCNRPEQEIRFTIKFQEFS 118

Qy   122 PNLWGHEFRSHHDYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKPV 181
               ||||| ||.||||||||||||| || ||||:|||||||||||||:|||:|||||||||
Db   119 PNYMGLEFKKHHDYYITSTSNGSLEGLENRGGVCRTRTMKIIMKVGQDPNAVTPEQLTT 178

Qy   182 SEMPMERDRGAAHSLE-PGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAGGLA---- 236
               || |.|.|.. |:| |||:|||||::|.|:| |: | ||:|.| ||.|| |
Db   179 SRPSKEADNTVKWATQAPGSRGSLGDSDGKHETVNQEEKSGP--------GASGGSSGDPD 231

Qy   237 ------LILLGVAGAGGA------MCWRRRRAKPSESRHPGPGSFGRGGSLGL 277
                      |||:| | || ||:| | |: |:||| |||| |||
Db   232 GFFNSKVALFAAVGAGCVIFLLIIFLTVLLLKLRKRHRKHTQQ-------RAAALSL 282

Qy   278 --GGGGGMGPREAEPGELGIALRGGGAADPFCPHYEKVSGDYGHPVYIYVQDGPPQSP 333
               | : ||.:|||||| : |||||||::.||||||||||||||||.: ||| |||
Db   283 STLASPKGGSGTAGTEPSDIIIPLR---TTENNYCPHYEKVSGDYGHPVYIYVQEMPPQSP 339

```
Qy   334 PNIYY 338
         ||||
Db   340 ANIYY 344
```

RESULT 12

```
ID   AAW19249           standard; Protein; 346 AA.
XX
AC   AAW19249;
XX
DT   18-AUG-1997  (first entry)
XX
DE   Human elk ligand protein.
XX
KW   Human; elk; ligand; elk-L; cytokine; testing; measuring;
KW   purification; neuroprotection; treatment; diabetic; hereditary;
KW   nutritional; neuropathy; neurodegenerative disease;
KW   tissue culture.
XX
OS   Homo sapiens.
XX
FH   Key            Location/Qualifiers
FT   Peptide        1..24
FT                  /label= sig_peptide
FT   Peptide        25..346
FT                  /label= mat_peptide
XX
PN   US5627267-A.
XX
PD   06-MAY-1997.
XX
PF   13-NOV-1992;  92US-0977693.
XX
PR   15-MAR-1994;  94US-0213403.
PR   13-NOV-1992;  92US-0977693.
PR   01-JUN-1995;  95US-0458077.
XX
PA   (IMMV ) IMMUNEX CORP.
XX
PI   Baum PR,  Beckmann MP,  Lyman S;
XX
PT   Human elk ligand protein - for diagnostic or therapeutic use, e.g.
PT   as neuro-protective agent
XX
PS   Claim 1; Columns 29-32; 18pp; English.
XX
CC   The present sequence is a human elk ligand (elk-L) protein,
CC   which binds elk, has a calculated molecular weight of 35180 and an
CC   isoelectric point of 9.006. Elk-L is a cytokine that can be used to
CC   test cells for elk expression, measure the biological activity of
CC   elk, purify elk by affinity chromatography and as a neuroprotective
CC   agent to treat diabetic, hereditary and nutritional neuropathies
CC   and neurodegenerative diseases. It may also be added to tissue
CC   cultures to prolong the life of neurons. The elk-L cDNA was
CC   isolated from a human placental cDNA library, and is present as a
CC   cDNA insert in the recombinant vector deposited in strain
CC   ATCC 69085.
XX
SQ   Sequence   346 AA;
```

Query Match   25.4%;   Score 623;   DB 18;   Length 346;
Best Local Similarity   39.2%;   Pred. No. 2.9e-44;
Matches 143;   Conservative 48;   Mismatches 116;   Indels 58;   Gaps 9;

```
Qy      8 PGGVRVGALILLGVLGLVSGL------------SLEPYVWNSANKRFQAEGGYVLYPQIGDRLDLL 61
          | :||| || :||:|: ||  :            |  |||:|||: :|  | : ||| ||  | |:
Db      4 PGQRWLGKWLVAMVVWALCRLATPLAKNLEPVSWSSLNPKFLSGKGLVIYPKIGKLDII 63

Qy     62 CPRARPPGPHSSPNYEFYKYLVGGAQGRRCEAPAPNLLLTCDRPDLDLRFTIKFQEYS 121
```

```
Db   340 ANIYY 344
```

RESULT 13

```
ID   AAW44323           standard; Protein; 346 AA.
XX
AC   AAW44323;
XX
DT   27-MAY-1998  (first entry)
XX
DE   Human elk-L.
XX
KW   Human; elk-L; elk ligand; cytokine; antibody; cell surface receptor;
KW   tyrosine kinase receptor; neural disease; trophic mechanism.
XX
OS   Homo sapiens.
XX
FH   Key            Location/Qualifiers
FT   Peptide        1..24
FT                  /label= signal
FT   Peptide        25..346
FT                  /label= elk-L
XX
PN   US5728813-A.
XX
PD   17-MAR-1998.
XX
PF   12-NOV-1996;  96US-0747240.
XX
PR   15-MAR-1994;  94US-0213403.
PR   13-NOV-1992;  92US-0977693.
PR   02-JUN-1995;  95US-0460741.
PR   12-NOV-1996;  96US-0747240.
XX
PA   (IMMV ) IMMUNEX CORP.
XX
PI   Baum PR,  Beckmann MP,  Lyman S;
XX
PT   Antibodies to elk ligand polypeptides - bind to tyrosine kinase
PT   receptors, useful for treating neural disease
XX
PS   Claim 1; Column 31-32; 17pp; English.
XX
DR   WPI; 1998-206621/18.
DR   N-PSDB; AAV15226.
XX
CC   The present sequence represents human elk-L (elk ligand). The present
CC   invention describes antibodies to elk-L polypeptides. The elk-L protein
CC   is known to be neuroprotective exhibiting its effect through a trophic
CC   mechanism. Examples of diseases that may be treated with elk-L include
CC   neuropathies e.g. diabetic, hereditary and nutritional neuropathies,
```

```
DR   WPI; 1997-271366/24.
DR   N-PSDB; AAT69766.
```

```
Qy     64 CPRARPAGRP--------YEYYKYLVRPEQAAACSTVLDPNVLVICNRPEQEIRFIIKFQEFS 118
             ||||     |:||||||     ||:||  | | :|:|| :  ||  | ||: :||||||||||:|
Db     64 PNLWGHEFRSHHDYYIIATSDGTREGLESSLQGGVCLIRGMKVLLLRVGQSPRGGAVPRKPV 118

Qy    119 PNTMGLEFKKHHDYYITSTSNGSLEGLENREGGVCRRTMKIIMRVGQDPNAVTPEQITT 178
          || :|||::| | : | | :| |||:|| | ||:|   : |  |  ||  | | :
Db    119 SRPSKEADNTVKMATQAPGSRGSLGDSSDGKHETVNQEBKSGP------GASGGSGSGDPD 178

Qy    179 SEMPMERDRGAAHSLE-PGKENLPGDPTSNATSRGAGGPLPPPSMPAVAGAAGGLA---- 236
          :|:|  :|   :| | |  |    | |   |:   |  || : | |:| | | |
Db    182 ----LLLLGVIAGAGGA------------MCWRRRAKPSESRHPGPGSFGRGGSLGL 236

Qy    237 ----GGGGGGMGPRBAEPGELGIALRGGGAAADPPFCPHYEKVSGDYGHPVIVQDGPPGSP 277
             |:: :||: ||||| |||:||::|||  |  |  |         | || :  :
Db    232 GFFNSKVALFAAVGAGCVIFLLIIIFLTVLLLKLRKRHRKHTQQ-------RAAALISI 282

Qy    278 ----MCWRRRAKPSESRHPGPGSFGRGGSLGL 277
Db    283 STLASPKGGSGTAGTEPSDIIIPLR--TTENNYCPHYEKVSGDYGHPVIVQEMPPGSP 339
```

```
Qy    334 PNIYY 338
          |||
Db    340 ANIYY 344
```

```
CC   neurodegenerative diseases and other diseases characterised by loss of
CC   function or degeneration of neurons.
XX
SQ   Sequence   346 AA;

Query Match             25.4%;  Score 623;  DB 19;  Length 346;
Best Local Similarity   39.2%;  Pred. No. 2.9e-44;
Matches  143;  Conservative   48;  Mismatches 116;  Indels  58;  Gaps   9;

Qy       8 PGGVRVGALILLGVIGLLVSGL---------SLEPVYWNSANKRFQAEGGYTVLYPQIGDRLDLL  61
                  |:|:|  |   |    |::|:||  |:  ||||| ||| |    ||:||||:
Db       4 PGQRWLGKWLVAMVWALCRLATPLAKNLEPVSWSSLNPKFLSGKGLVIYPKIGDKLDII  63

Qy      62 CPRARPPGPHSSPNYEYKLYLVGGAQGRRCEAPPAPENLLLTCDRPDLDLRFTIKFQEYS 121
            |||| ||        ||:||||||| | |::  | | |  :|||||:|||||||||:
Db      64 CPRAEAGRP------YEYKLYLVRPEQAAACSTVLDPNVLVTCNRPEQEIRFTIKFQEFS 118

Qy     122 PNIMGHEFRSHHDYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKPV 181
            |: :|||||||||||||||||||||||||||||||||||||||||||||:
Db     119 PNYMGLEFKKHHDYYITSTNGSLBGLENREGGVCRIRTMKIIMKVGQDPNAVTPEQLTT 178

Qy     182 SEMPMERDRGAAHSLE-PGKENLPGDPFTSNATSRGAEGPLPPPSMPAVAGAAGGLA---- 236
            ::  ||||||||::|| |||||||||   |: |  |  |  |  |||:|  |   ||
Db     179 SRPEKEADNTVKNATQAPGSRGSLGDSDGKHETVNQEEKSGP-----GASGGSSGDPD 231

Qy     237 ------LLILGVAGAGGA--------------MCWRRRRAKPSESRHPGPGSFGRGGSLGL 277
                  ||||||||||              |||||||||||||||||||||||:
Db     232 GFFNSKVALFAAVGAGCVIFLLIIIFLTVLLLKLRKRHRKHTQQ---------RAAALSI 282

Qy     278 ---GGGGGMGPREAEPGELGIALRGGGAADPPCPHYEKVSGDYGHPVIVIVQDGPPEQSP 333
              |||||||||||||||||||||||||||||||||||||||||||||||:
Db     283 STLASPKGGSGTAGTEPSDIIIPLR---TTENNYCPHYEKVSGDYGHPVYIVQEMPPQSP 339

Qy     334 PNIIY 338
           ||||
Db     340 ANIYY 344
```

```
Query Match             25.4%;  Score 622;  DB 16;  Length 346;
Best Local Similarity   40.5%;  Pred. No. 3.5e-44;
Matches  137;  Conservative   44;  Mismatches 105;  Indels  52;  Gaps   8;

Qy      29 SLEPVYWNSANKRFQAEGGYTVLYPQIGDRLDLLCPRARPPGPHSSPNYEYKLYLVGGAQ  88
           |:||||| ||| |    ||:|||::|  ||:|||||||||| ||        ||:|||||||
Db      31 NLEPVSWSSLNPKFLSGKGLVIYPKIGDKLDIICPRAEAGRP------YEYKLYLVRPEQ  85

Qy      89 GRRCEAPPAPENLLLTCDRPDLDLRFTIKFQEYSPNIMGHEFRSHHDYIIATSDGTREGL 148
            |::  | | |  :|||||:|||||||||:|||:|  ||||||||||||||
Db      86 AAACSTVLDPNVLVTCNRPEQEIRFTIKFQEFSPNYMGLEFKKHHDYYITSTNGSLEGL  145

Qy     149 ESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKPVSEMPMERDRGAAHSLE-PGKENLPGDP 207
            ||||||||||||||||||||||||:          ::  |||||||::|| |||||||||
Db     146 ENREGGVCRIRTMKIIMKVGQDPNAVTPEQLTTSRPEKEADNTVKNATQAPGSRGSLGDS 205

Qy     208 TSNATSRGAEGPLPPPSMPAVAGAAGGLA--------------LLILGVAGAGGA------ 248
                 |  |  |  |  |||:|  |   ||              ||||||||||
Db     206 DGKHETVNQEEKSGP-----GASGGSSGDPDGFFNSKVALFAAVGAGCVIFLLIIIFL 258

Qy     249 -----MCWRRRRAKPSESRHPGPGSFGRGGSLGL----GGGGGMGPREAEPGELGIALRGG 300
                 |||||||||||||||||||:          |||||||||||||||||||||||
Db     259 TVLLLKLRKRHRKHTQQ---------RAAALSLSTLASPKGGSGTAGTEPSDIIIPLR--  307

Qy     301 GAADPPCPHYEKVSGDYGHPVIVIVQDGPPEQSPPNIIY 338
            :||||||||||||||||||||:          ||||
Db     308 -TTENNYCPHYEKVSGDYGHPVYIVQEMPPQSPANIYY 344
```

RESULT 16

AAW06337

```
ID   AAW06337 standard; Protein; 333 AA.
XX
AC   AAW06337;
XX
DT   17-JAN-1997 (first entry)
XX
DE   Full length ligand for receptor-type tyrosine kinase protein.
XX
KW   Receptor-type tyrosine kinase ligand; coomassie staining; PAS staining;
KW   human.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..25
XX
PR   27-FEB-1995; 95US-0395415.
XX
PA   (HARD ) HARVARD COLLEGE.
XX
PI   Bergemann AD, Flanagan JG;
XX
DR   WPI; 1996-433391/43.
DR   N-PSDB; AAT40231.
XX
PT   Eph receptor ligand, ELF-2, and DNA encoding it - used to treat or
PT   prevent neurological diseases, and to modulate binding of ELF-2 to
PR   Eph receptor, e.g. to prevent or treat tumour formation.
XX
PS   Claim 6; Fig 2A-D; 50pp; English.
XX
CC   Chicken Eph receptor ligand ELF-2 (AAW00288) is a ligand important
CC   in cellular communication during pattern formation. It is strongly
CC   expressed in the anterior hindbrain and newly-forming somites of
CC   embryos at the early organogenesis stage of development. Its amino
CC   acid sequence was deduced from a cDNA clone (AAT40231) isolated from
CC   a chicken cDNA library. The ELF-2 ligands can be used to alter
CC   neurological development, oncogenesis and growth regulation, to
CC   modulate binding of ELF-2 to the Eph receptor, and in diagnostic
CC   assays.
XX
SQ   Sequence 331 AA;

Query Match          25.4%;  Score 621.5;  DB 17;  Length 331;
Best Local Similarity  40.2%;  Pred. No. 3.7e-44;
Matches  133;  Conservative  53;  Mismatches 124;  Indels  21;  Gaps  6;

Qy  14 GALLLLGVLGLVSGLSLEPVYWNSANKRFQAEGGGYVLYPQIGDRLDLLCPRARPPGHSS 73
        ||||:||||| | |:|||:||::|| |||||||||||:|::|||||||||||||
Db  14 GALMVLWRTALAKSIVLDPIYWNSSNPKFLPGQGLVLVYPQIGDKLDIICPKV---DSKTA 70

Qy  74 PNYEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDLDLRFTIKFQEYSPNLWGHEFRSHH 133
        ||:|:|||:| | |:||||:||||||||||:|||::|||||||||:|||||||:||:.
Db  71 GLYEYYKVYMVDKDQADSCAIRKDNTPLLNCAKPDQDVKFTIKFQEFSPNLWGLEFQKNK 130

Qy 134 DYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKPVSEMPMER----- 188
        ||||::||:|:|||:.|:|||||.|:.|::|:|||:|...::|.|||||||:|
Db 131 DYYVISTSNGSLEGLNNQEGGVCQTKTMILMKVGQDPNSAGLPER---SSDPTKRPEQEA 187

Qy 189 -DRGAAHSLEPGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAGGLALLLLGVAGAGG 247
         :..||:||.|:.| .||:|:|:.||..|.|:..::||||
Db 188 GTNGXSSTTSPFVKDHSGSSTDG---SKAGHSSILGSEVALFAGIASGCIIFIVIIILVV 245

Qy 248 AMCWRRRRAKPSESSRHPGPGSFGRGGSLGLGGGGGMGPREAEPGELGIALRGGGAADPPF 307
        :::.|||.|.|:.:.||.:...:..:.:|.::..||.
Db 246 LILIXYRRHRKHSPQHTTTLSLSTLATPKRSGNN----NGSEFSDIIIPLR--TADSVF 298

Qy 308 CPHYEKVSGDYGHPVYIVQDGPPQSPPNIYY 338
        |||||||||||||||||||:.|||||.|||
Db 299 CPHYEKVSGDYGHPVYIVQEMPPQSPANIYY 329
```

```
Protein          /note= "signal peptide"
                 /note= 26..333
                 /note= "receptor-type tyrosine kinase binding protein"
XX
FN   JP08188596-A.
XX
PD   23-JUL-1996.
XX
PF   13-JAN-1995; 95JP-0003677.
XX
PR   09-NOV-1994; 94JP-0275411.
PR   19-OCT-1994; 94JP-0253848.
XX
PA   (ASAH ) ASAHI KASEI KOGYO KK.
XX
DR   WPI; 1996-388601/39.
DR   N-PSDB; AAT42594.
XX
PT   New ligand for receptor type tyrosine kinase - has mol.wt. 22-25
PT   kilo:dalton(s) and is positive for Coomassie and PAS staining
XX
PS   Example 11; Page 44-46; 51pp; Japanese.
XX
CC   This sequence represents the full length receptor-type tyrosine kinase
CC   protein binding ligand of the invention (also see AAW06333 and
CC   AAW06334). The mature protein contained within this sequence contains the
CC   N-terminal fragment represented by AAW06332. The proteins of the
CC   invention have a molecular weight of 23500 (plus or minus 1500) Da, and
CC   are positive for Coomassie staining and PAS staining. The proteins of
CC   the invention bind to the fragment of this protein sequence represented
CC   by AAW06330. The proteins of the invention are new ligands of
CC   receptor-type tyrosine kinases, and can be prepared by standard
CC   recombinant techniques.
XX
SQ   Sequence 333 AA;

Query Match          25.3%;  Score 620.5;  DB 17;  Length 333;
Best Local Similarity  40.5%;  Pred. No. 4.5e-44;
Matches  133;  Conservative  52;  Mismatches 130;  Indels  13;  Gaps  5;

Qy  14 GALLLLGVLGLVSGLSLEPVYWNSANKRFQAEGGGYVLYPQIGDRLDLLCPRARPPGHSS 73
        ||||:|||:.|:.|:|||:||::||.|||||||||||:|::|||||||||||||
Db  14 GVLMVLCRTAISKSIVLEPIYWNSSNSKFLPGQGLVLVYPQIGDKLDIICPKV---DSKTV 70

Qy  74 PNYEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDLDLRFTIKFQEYSPNLWGHEFRSHH 133
        ||:|:|||:| | |:||||:|.::|||:|.||||::|.|||||:|||||||:||:.
Db  71 GQYEYYKVYMVDKDQADRCTIKKENTPLINCAKPDQDIKFIKFQEFSPNLWGLEFQKNK 130

Qy 134 DYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQ-SPRGGAVPRKPVSEMPMER--DR 190
        ||||::|:|.:|||:.|:|||||.|..|::|:|||:|...::..|:.|||.|
Db 131 DYYIISTNGSLEGLDNQEGGVCQTRAMKILMKVGQDPASSAGSTRNKDPTRRPELEAGTN 190

Qy 191 GAAHSLEPGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAGGLALLLLGVAGAGAMC 250
        |:..||.|.:.| .|:|:|:.||.:.|.|:..::||||
Db 191 GRSSTTSPFVKPNPGSSTDGNSAGHSGNNILGSEVALFAGIASGCIIFIVIIILVLLL 250

Qy 251 WRRRRAKPSESSRHPGPGSFGRGGSLGLGGGGGMGPREAEPGELGIALRGGGAADPPPCPH 310
        .|||.|.|:.:.||.:...:..:.:|.::..||.
Db 251 KYRRHRKHSPQHTTTLSLSTLATPKRSGNN----NGSEFSDIIIPLR---TADSVFCPH 303

Qy 311 YEKVSGDYGHPVYIVQDGPPQSPPNIYY 338
        |||||||||||||||:.|||||.|||
Db 304 YEKVSGDYGHPVYIVQEMPPQSPANIYY 331
```

RESULT 17

AAR94655

```
ID   AAR94655 standard; Protein; 333 AA.
XX
AC   AAR94655;
XX
DT   18-OCT-1996 (first entry)
XX
```

```
RESULT 18
AAR92743
ID   AAR92743 standard; Protein;  333 AA.
```

Left column

```
DE   Ligand for receptor type tyrosine kinase (TK).
XX
KW   Receptor type tyrosine kinase; TK; ligand; differentiation;
KW   haematopoietic stem cell; tyrosine; bone marrow; leukaemia.
XX
OS   Homo sapiens.
XX
PN   WO9611212-A1.
XX
PD   18-APR-1996.
XX
PF   09-OCT-1995;   95WO-JP02069.
XX
PR   22-DEC-1994;   94JP-0320712.
PR   07-OCT-1994;   94JP-0244433.
PR   26-OCT-1994;   94JP-0262882.
XX
PA   (ASAHI ) ASAHI KASEI KOGYO KK.
XX
PI   Ohno M,  Sakano S;
XX
DR   WPI; 1996-208809/21.
DR   N-PSDB; AAT18395.
XX
PT   Ligand peptide binding to receptor-type tyrosine kinase - enhances
PT   intracellular tyrosine phosphorylation, useful for investigation of
PT   undifferentiated blood cell behaviour
XX
PS   Claim 6; Page 163-164; 193pp; Japanese.
XX
CC   A ligand polypeptide which binds to the extracellular part of a
CC   specific receptor-type tyrosine kinase and induces phosphorylation
CC   of tyrosine within the cell can be used in the study of the
CC   differentiation of blood cells such as the haematopoietic stem
CC   cells, of disease processes such as leukaemia, and of the biology of
CC   bone marrow transplantation. The ligand plays a role in the
CC   differentiation process and the specific ligand target is expressed
CC   in undifferentiated blood cells.
XX
SQ   Sequence   333 AA;
XX

Query Match    25.3%;   Score 620.5;   DB 17;   Length 333;
Best Local Similarity   40.5%;   Pred. No. 4.5e-44;
Matches  133;   Conservative   52;   Mismatches 130;   Indels  13;   Gaps   5;

Qy      14 GALLLLGVLGLVSGLSLEPVYWNSANKRFQAEGGYVYLYPQIGDRLDLLCPRAPPGPHSS 73
           |:|| |  |::|  |:|
Db      14 GVLMVLCRTAISKSIVLEPIYWNSSNSKFLPGQGLVLYPQIGDKLDIICPKV---DSKTV 70

Qy      74 PNYEFYKLYLVGGAQGRRCEAPPAPNLLLTCRPDLDLRFTIKFQEYSPNLWGHEFRSHH 133
           |:||:|||:|::|||::: :| |:||| |  |||:|:|:|||  ||||:|||||:|:
Db      71 GQYEYYKVYMVDKDQADRCTIKKENTPLLNCAKPDQDIKFTIKPQEFSPNLWGLEFQKNK 130

Qy     134 DYYIIATSDGTRGLESLQGGVCLTRGMKVLLRVGQ--SPRGGAVPRKPVSEMPMER-DR 190
           |::::|| |  |:|:|||::|||:|||||:||:||    |:|||| ||||: |:|   |
Db     131 DYYIISTSNGSLEGLDNQEGGVCQTRAMKILMKVGQDASSAGSTRNKDPTRRPELEAGTN 190

Qy     191 GRSSTTSPFVKPNPGSSTDGNSAGHSGNNILGSEVALFAGIASGCIFIVIILVVLLL 250
           |  |  |::|    |:|     |   |   |||::||  |||||  |:| :|| |||:
Db     191 GAAHSLEPGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAGGLALLLLGVAGAGGAMC 250

Qy     251 WRRRAKPSESRHPGPGSPGRGGSLGLGGGGMGPRAEPGELGIALRGGGAADPPPCPH 310
           | | ||||  |   |:|   |  |  |     ||    ||:  | |  |  | ||||
Db     251 KYRRRHRKHSPQHTTTLSLSTLATPKRSGNN---NGSEPSDIIIPLR---TADSVPCPH 303

Qy     311 YEKVSGDYGHPVYIVQDGPPQSPENIYY 338
           |||:|||:|||||:|:|:|| |:||: |
Db     304 YEKVSGDYGHPVVIVQEMPPQSPANIYY 331
```

Right column

```
XX
AC   AAR92743;
XX
DT   21-MAY-1996  (first entry)
XX
DE   Human hepatoma transmembrane kinase receptor ligand.
XX
KW   Hepatoma transmembrane kinase; Htk; receptor; ligand;
KW   tyrosine kinase; ligand; neurodegenerative disease.
XX
OS   Homo sapiens.
XX
PN   WO9602645-A2.
XX
PD   01-FEB-1996.
XX
PF   14-JUL-1995;   95WO-US08812.
XX
PR   20-JUL-1994;   94US-0277722.
XX
DR   WPI; 1996-105907/11.
DR   N-PSDB; AAT16471.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Bennett BD,  Matthews W;
XX
PT   Ligand for the hepatoma trans-membrane kinase receptor - useful for
PT   stimulating and inhibiting cells carrying the receptor, e.g. for
PT   treating neuro-degenerative disease
XX
PS   Claim 6; Fig 2(A-B);  88pp; English.
XX
CC   Mouse (AAT16470) and human (AAT16471) Htk ligand which bind to, and
CC   activate, the Htk receptor, have been identified in a variety of
CC   tissues using a soluble Htk-Fc fusion protein.
CC   The predicted mol.wt. of the murine Htk ligand protein following a
CC   signal peptide cleavage is 34 kD with an estimated pI of 8.9.
CC   The murine and human ligands show 96% homology at the amino acid
CC   level.
CC   The DNA is used to produce recombinant ligands; for tissue-
CC   specific typing (partic. as a marker for breast cancer) and as a
CC   marker for human chromosome 13. The ligands (partic. in soluble
CC   form) are used to activate the tyrosine kinase domain of the
CC   Htk receptor, i.e. to stimulate or inhibit growth, differentiation,
CC   and/or activation of cells contg. the receptor, e.g. treatment
CC   of neurodegenerative diseases, since they are strongly expressed
CC   in the cerebral cortex, hippocampus, striatum and cerebellum.
CC   The ligands are also useful as a control or standard in assays,
CC   for generation of antibodies, as a mol. wt. marker for growth
CC   in vitro of Htk-receptor positive cells, as research agent,
CC   in screening, etc.
XX
SQ   Sequence   333 AA;
XX

Query Match    25.3%;   Score 620.5;   DB 17;   Length 333;
Best Local Similarity   40.5%;   Pred. No. 4.5e-44;
Matches  133;   Conservative   52;   Mismatches 130;   Indels  13;   Gaps   5;

Qy      14 GALLLLGVLGLVSGLSLEPVYWNSANKRFQAEGGYVYLYPQIGDRLDLLCPRAPPGPHSS 73
           |:|| |  |::|  |:|
Db      14 GVLMVLCRTAISKSIVLEPIYWNSSNSKFLPGQGLVLYPQIGDKLDIICPKV---DSKTV 70

Qy      74 PNYEFYKLYLVGGAQGRRCEAPPAPNLLLTCRPDLDLRFTIKFQEYSPNLWGHEFRSHH 133
           |:||:|||:|::|||::: :| |:||| |  |||:|:|:|||  ||||:|||||:|:
Db      71 GQYEYYKVYMVDKDQADRCTIKKENTPLLNCAKPDQDIKFTIKPQEFSPNLWGLEFQKNK 130

Qy     134 DYYIIATSDGTRGLESLQGGVCLTRGMKVLLRVGQ--SPRGGAVPRKPVSEMPMER-DR 190
           |::::|| |  |:|:|||::|||:|||||:||:||    |:|||| ||||: |:|   |
Db     131 DYYIISTSNGSLEGLDNQEGGVCQTRAMKILMKVGQDASSAGSTRNKDPTRRPELEAGTN 190

Qy     191 GAAHSLEPGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAGGLALLLLGVAGAGGAMC 250
```

Db   191 GRSSTTSPFVKPNPGSSTDGNSAGHSGNNILGSEVALFAGIASGCIIFIVIILLVVLLL 250
XX
Qy   251 WRRRRAKPSESRHPGPGSFGRGGSLGLGGGGMGPREAEPGELGIALRGGGAADPPFCPH 310
          |: :| |:. :|| ||||||||||||||||||||
Db   251 KYRRRHRKHSPQHTTTLSLSTLATPKRSGNN---NGSEPSDIIIPLR---TADSVFCPH 303
XX
Qy   311 YEKVSGDYGHPVYIVQDGPPQSPPNIYY 338
          ||||||||||||||:||:| :|| |||
Db   304 YEKVSGDYGHPVYIVQEMPPQSPANIYY 331

RESULT 19
ID   AAR89287   standard; Protein; 333 AA.
XX
AC   AAR89287;
XX
DT   13-APR-1996   (first entry)
XX
DE   Human LERK-5 cytokine.
XX
KW   Human; LERK-5; cytokine; elk; hek; protein-tyrosine-kinase;
KW   receptor; chromosome-13; expressed sequence tag; T-lymphocyte;
KW   fetal brain; cloning; soluble LERK-5; antibody; Fc; fusion protein;
KW   dimer; FLAG; affinity tail; drug delivery; cell targeting;
KW   cytostatic; leukaemia; therapy.
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FH
FT   Peptide   1..25
FT   /note= "Signal peptide"
FT   Protein   1..224
FT   /note= "Soluble LERK-5 (claim 3)"
FT   Peptide   26..37
FT   /note= "N-terminal peptide (claim 15)"
FT   Domain   26..224
FT   /note= "Extracellular domain (mature soluble
FT   LERK-5, claim 3)"
FT   Modified-site   36..38
FT   /note= "N-glycosylation site"
FT   Region   62
FT   /note= "Conserved cysteine residue"
FT   Region   89
FT   /note= "Conserved cysteine residue"
FT   Region   101
FT   /note= "Conserved cysteine residue"
FT   Region   104..215
FT   /note= "Conserved cysteine residue"
FT   Region   139..141
FT   /note= "Chromosome-13 expressed sequence tag
FT   GenBank L13189 translation"
FT   Modified-site   153
FT   /note= "N-glycosylation site"
FT   Region   225..250
FT   /note= "Conserved cysteine residue"
FT   Region   225..250
FT   /note= "Transmembrane region"
FT   Domain   251..333
FT   /note= "Cytoplasmic domain"
FT   Cleavage-site   253..254
FT   /note= "KEX2 protease processing site"
FT   Cleavage-site   254..255
FT   /note= "KEX2 protease processing site"
FT   Cleavage-site   257..258
FT   /note= "KEX2 protease processing site"
FT   Cleavage-site   276..277
FT   /note= "KEX2 protease processing site"
PN   WO9601839-A1.
XX
PD   25-JAN-1996.
XX
PF   06-JUL-1995;   95WO-US08534.
XX

PR   08-JUL-1994;   94US-0271948.
XX
PA   (IMMV) IMMUNEX CORP.
XX
PI   Cerretti DP, Reddy P;
XX
DR   WPI; 1996-097585/10.
DR   N-PSDB; AAQ99288.
XX
PT   DNA encoding LERK-5 cytokine capable of binding elk and hek - used
PT   as carrier to deliver attached (anti-leukaemic) agents to cells
PT   bearing elk or hek, i.e. human leukaemia cell lines JM and LK63
XX
PS   Claim 1; Page 31-32; 40pp; English.
XX
CC   The sequence represents a full-length LERK-5 cytokine, which binds
CC   elk and hek receptor protein-tyrosine-kinases (members of the
CC   eph/elk family), and expressed from a gene isolated by amplification
CC   with primers specific for a chromosome-13 expressed sequence tag,
CC   GenBank L13189; from a human T-lymphocyte leukaemia CCRF-HSB-2 (ATCC
CC   CCL 120.1) cell line, followed by cDNA probe screening of a human
CC   fetal brain cDNA library in phage lambda-gt10. A truncated soluble
CC   form may also be produced. The protein may be produced as a
CC   recombinant fusion protein or dimer with an antibody Fc fragment or
CC   e.g. a FLAG affinity tail (R89288). LERK-5 may be used in binding
CC   assays to detect elk- or hek-expressing cells, to measure elk or hek
CC   activity, as a drug delivery for elk or hek receptor-specific cell
CC   targeting, e.g. for delivery of cytostatic agents in leukaemia
CC   therapy.
CC
XX
SQ   Sequence   333 AA;

Query Match   25.3%;   Score 620.5; DB 17;   Length 333;
Best Local Similarity   40.5%;   Pred. No. 4.5e-44;
Matches 133; Conservative 52; Mismatches 130; Indels 13; Gaps 5;

Qy   14 GALILILGVLGLVSGLSLBPVYWNSANKRFQABGGYVLYPQIGDRLDLLCPRARPGPHSS 73
          |::| |.|:|:::.:||::: |:||||
Db   14 GVLMVLCRTAISKSIVLEPIYWNSSNSKFLPGQGLVLYPQIGDKLDIICPKV---DSKTV 70

Qy   74 PNYEPYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDLDLRFTIKFQEYSPNLWGHEFRSHH 133
          |:||| .|: |:.|
Db   71 GQYEYKVYMVDKDQADRCTIKKENTPLLNCAKPDQDIKFTIKFQEFSPNLWGLBFQKNK 130

Qy   134 DYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQ--SPRGGAVPRKPVSEMMER-DR 190
          .|||| ||| |:||:   :||:|:||
Db   131 DYYIISTNGSLRELGLDNQEGGVCQTRAMKILMKVGQDASSAGSTRNKDPTRRPELEAGTN 190

Qy   191 GAARSLEPGKENLPGDPTPSNATSRGAEGPLPPPSMPAVAGAGAGGLALLLLGLVGAGAGC 250
          |: |:||::| : ::|| :: |: |||:|||: ||:|||||||:
Db   191 GRSSTTSPFVKPNPGSSTDGNSAGHSGNNILGSEVALFAGIASGCIIFIVIILLVVLLL 250

Qy   251 WRRRRAKPSESRHPGPGSFGRGGSLGLGGGGMGPREAEPGELGIALRGGGAADPPFCPH 310
          |: :| |:. :|| ||||||||||||||||||||
Db   251 KYRRRHRKHSPQHTTTLSLSTLATPKRSGNN---NGSEPSDIIIPLR---TADSVFCPH 303

Qy   311 YEKVSGDYGHPVYIVQDGPPQSPPNIYY 338
          ||||||||||||||:||:| :|| |||
Db   304 YEKVSGDYGHPVYIVQEMPPQSPANIYY 331

RESULT 20
ID   AAW11308   standard; protein; 333 AA.
XX
AC   AAW11308;
XX
DT   21-MAR-1997   (first entry)
XX
DE   Receptor-type tyrosine kinase #4.
XX
KW   Receptor-type tyrosine kinase; LPM medium; ligand; autophosphorylation;
KW   insect cell; animal cell; growth promoter; undifferentiated cell;

Left column (RESULT 21)

```
KW   insulin; interleukin; fibroblast growth factor; hepatocyte growth factor;
KW   nerve growth factor; interferon-gamma; tumour necrosis factor; inducer.
XX
OS   Homo sapiens.
XX
FH   Key          Location/Qualifiers
FT
FT   Peptide      1..25
FT                /label= signal peptide
FT   Protein      26..333
FT                /label= receptor-type tyrosine kinase
XX
PN   JP08280385-A.
XX
PD   29-OCT-1996.
XX
PF   13-APR-1995;  95JP-0087878.
XX
PR   13-APR-1995;  95JP-0087878.
XX
PA   (ASAHI ) ASAHI KASEI KOGYO KK.
XX
PR
XX
DR   WPI; 1997-014848/02.
DR   N-PSDB; AAT51236.
XX
CC   AAW1307 and AAW1308 represent receptor-type tyrosine kinases. These
CC   sequences are included in the LPM medium of the invention. The medium
CC   is used to culture cells capable of producing a ligand for receptor-type
CC   tyrosine kinase. The ligand serves as a physiologically active
CC   substance to induce enzyme activity and autophosphorylation of tyrosine
CC   kinase, which is involved in the development and differentiation of
CC   animal and insect cells. The ligand can also be used as in a
CC   pharmaceutical preparation for promoting growth of undifferentiated
CC   cells. The medium may also include an inducer, such as insulin, an
CC   interleukin, a fibroblast growth factor, or interferon-gamma.
XX
PT   LPM cell culture medium  - for culturing cells which produce
PT   receptor-type tyrosine kinase ligand
PS   Disclosure; Page 40-41; 45pp; Japanese.
XX
SQ   Sequence  333 AA;

Query Match   25.3%;
Best Local Similarity   40.5%;
Matches 133;  Conservative 52;  Mismatches 130;  Indels 13;  Gaps 5;

Qy   14 GALLLLGVLGLVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLLDLLCPRARPGPHSS 73
         |:||                   ||||||||||||||:||:||:
Db   14 GVLMVLCRTAISKSIVLLEPIYWNSSNSKFLPGQGLVLYPQIGDKLDIICPKV---DSKTV 70

Qy   74 PNYEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDLDLRFTIKFQEYSPNLWGHEFRSHH 133
         ||:||||             :|:|||||||||:||::||||||||||||||:
Db   71 GQYEYKVYMVDKDQADRCTIKKENTPLLNCAKPDQDIKFTIKFQEFSPNLWGLEFQOKNK 130

Qy  134 DYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQ--SPRGGAVPRKPVSEMPMER-DR 190
Db  131 DYYIISTNGSLEGLDNGEGGVCQTRAMKILMKVGQDASSAGSTRNKDPTRPELEAGTN 190

Qy  191 GAAHSLEPGKENIPGDPTPSNATSRGAEGPLPPSMPAVAGAAGGLALLLLGVAGAGAGAMC 250
Db  191 GRSSTTSPFVKPNPGSSTDGNSAGHSGNNILGSEVALFAGIASGCIIFIVIIITLLVLLL 250

Qy  251 WRRRAKPSESRIHPGPGSFGRGGSIGLGGGGMGPRRAEPGELGIALRGGGAADPPFCPH 310
Db  251 KYRRHRKHSPQHTTLSLSTLATPKRSGNN---NGSEPSDIIIPLR---TADSVFCPH 303

Qy  311 YEKVSGDYGHPVVIVQDGPPQSPPNIIY 338
Db  304 YEKVSGDYGHPVVIVQEMPPQSPANIYY 331

RESULT 21
AAW06334:
```

Right column (RESULT 22)

```
ID   AAW06334 standard; protein; 308 AA.
XX
AC   AAW06334;
XX
DT   17-JAN-1997  (first entry)
XX
DE   Ligand #2 for receptor-type tyrosine kinase protein.
XX
KW   Receptor-type tyrosine kinase; ligand; coomassie staining; PAS staining;
KW   human.
XX
OS   Homo sapiens.
XX
PN   JP08188596-A.
XX
PD   23-JUL-1996.
XX
PF   13-JAN-1995;  95JP-0003677.
XX
PR   09-NOV-1994;  94JP-0275411.
PR   19-OCT-1994;  94JP-0253848.
XX
PA   (ASAH ) ASAHI KASEI KOGYO KK.
XX
DR   WPI; 1996-388601/39.
XX
PS   Claim 7; Page 44; 5pp; Japanese.
PT   New ligand for receptor type tyrosine kinase  - has mol.wt. 22-25
PT   kilo:dalton(s) and is positive for Coomassie and PAS staining
XX
CC   AAW06333 and AAW06334 represent receptor-type tyrosine kinase receptor
CC   binding ligands of the invention. These sequences both have the
CC   N-terminal sequence represented by AAW06332. The ligands of the
CC   invention recognise the fragment of the receptor type kinase receptor
CC   represented by AAW06330 (see AAW06331 for full length sequence). The
CC   proteins of the invention have a molecular weight of 23500 (plus or minus
CC   1500) Da, and are positive for Coomassie staining and PAS staining. The
CC   protein is a new ligand of receptor-type tyrosine kinases, and can be
CC   prepared by standard recombinant techniques.
XX
SQ   Sequence  308 AA;

Query Match   24.9%;
Best Local Similarity   41.7%;
Matches 130;  Conservative 48;  Mismatches 121;  Indels 13;  Gaps 5;

Qy   30 LEPVYWNSANKRFQAEGGYVLYPQIGDRLLDLLCPRARPGPHSSSPNYEFYKLYLVGGAQG 89
         ||:||||||      ||||||||||:||:||:               :|:
Db    5 LEPIYWNSSNSKFLPGQGLVLYPQIGDKLDIICPKV---DSKTVGQYEYKVYMVDKDQA 61

Qy   90 RRCEAPPAPNLLLTCDRPDLDLRFTIKFQEYSPNLWGHEFRSHHDYYIIATSDGTREGLE 149
         |||     :|:||||||||:||::||||||||||||||:          ||||||||:
Db   62 DRCTIKKENTPLLNCAKPDQDIKFTIKFQEFSPNLWGLEFQOKNKDYYIISTNGSLEGLD 121

Qy  150 SLQGGVCLTRGMKVLLRVGQ--SPRGGAVPRKPVSEMPMER-DRGAAHSLEPGKENIPGD 206
Db  122 NQEGGVCQTRAMKILMKVGQDASSAGSTRNKDPTRPELEAGTNGRSSTTSPFVKPNPGS 181

Qy  207 PTSNATSRGAEGPLPPSMPAVAGAAGGLALLLLGVAGAGAGAMCWRRRAKPSESRHPGP 266
Db  182 STDGNSAGHSGNNILGSEVALFAGIASGCIIFIVIIITLLVLLLKYRRHRKHSPQHTTT 241

Qy  267 GSFGRGGSLGLGGGGMGPRRAEPGELGIALRGGGAADPPFCPHYEKVSGDYGHPVYIVQ 326
Db  242 LSLSTLATPKRSGNN----NGSEPSDIIIPLR--TADSVFCPHYEKVSGDYGHPVVIVQ 294

Qy  327 DGPPQSPPNIIY 338
Db  295 EMPPQSPANIYY 306

RESULT 22
```

```
ID   AAR94656 standard; Protein; 308 AA.
XX
AC   AAR94656;
XX
DT   18-OCT-1996  (first entry)
XX
DE   Receptor type tyrosine type tyrosine kinase (TK) (mature protein).
XX
KW   Ligand for receptor type tyrosine kinase; TK; ligand; differentiation;
KW   haematopoietic stem cell; tyrosine; bone marrow; leukaemia.
XX
OS   Homo sapiens.
XX
PN   WO9611212-A1.
XX
PD   18-APR-1996.
XX
PF   09-OCT-1995;  95WO-JP02069.
XX
PR   22-DEC-1994;  94JP-0320712.
PR   07-OCT-1994;  94JP-0244433.
PR   26-OCT-1994;  94JP-0262882.
XX
PA   (ASAH ) ASAHI KASEI KOGYO KK.
XX
PI   Ohno M, Sakano S;
XX
DR   WPI; 1996-209809/21.
DR   N-PSDB; AAT18395.
XX
PT   Ligand peptide binding to receptor-type tyrosine kinase - enhances
PT   intracellular tyrosine phosphorylation, useful for investigation of
PT   undifferentiated blood cell behaviour.
XX
PS   Disclosure; Page 162-163; 193pp; Japanese.
XX
CC   A ligand polypeptide which binds to the extracellular part of a
CC   specific receptor-type tyrosine kinase and induces phosphorylation
CC   of tyrosine within the cell can be used in the study of the
CC   differentiation of blood cells such as the haematopoietic stem
CC   cells; of disease processes such as leukaemia, and of the biology of
CC   bone marrow transplantation.  The ligand plays a role in the
CC   differentiation process and the specific ligand target is expressed
CC   in undifferentiated blood cells.
XX
SQ   Sequence   308 AA;

Query Match             24.9%;  Score 610.5;  DB 17;  Length 308;
Best Local Similarity   41.7%;  Pred. No. 2.8e-43;
Matches 130;  Conservative  48;  Mismatches 121;  Indels  13;  Gaps    5;

Qy     30 LEPVYWNSANRFQABGGYVYLYPQIGDRLDLLCPRARPGPHSSPNYEFYKLYLVGGAQG 89
          ||:|:|||:|||||||:|::||::|||||:|||||||:|||:|:|||:|||||:||||
Db      5 LEPIYWNSSNSKFLPGGLVLYPQLGDKLDIICPKV---DSKTVGQYEYKVTWVDKDQA 61

Qy     90 RRCEAPPAPNLLLTCDRPDLDLRFTIKFQEYSPNLWGHEFRSHHDYYIIATSDGTREGLE 149
          |||:||:||:|||::|||||:|:||:|:||||||||||:||:|:||:||::|:|||:||
Db     62 DRCITKKENTPLLNCAKPDQDIKFIKFQEFSPNIWGLEFQKNDYYIISTSNGSLEGLD 121

Qy    150 SLQGGVCLTRPGMKYLLRVGQ---SPRGGAVPRKPVSEMPMER-DRGAAHSLEPGKENLPGD 206
          |:||::|||:||||||:||||   :|||:||:||::|:||::||::||:|:|||:||||
Db    122 NQEGGVCQTRAMKLLMKVGQDASAGSTRNKDPTRPELEAGTNGRSSTTSPFVXPNPGS 181

Qy    207 PTSNATSRGAEGPLPPPSMPAVAGANAGGLALILLGVAGAGGAMCWRRRAKPSESRHPGP 266
          ::||:||:|::|:||:|:|||||||||||:|||:||:|::|:|:|::|:|::|:||||
Db    182 STDGNSAGHSGNNILGSEVALFAGISAGCIFIVTIIVVLLKYRRHRKHSPQHTTT 241

Qy    267 GSFGRGGSLGLGGGGMGRPEAEFGELGIALRGGGAADPPFCPHYEKVSGDYGHPVYIVQ 326
          |::|:||||:|:|||||||||::||:||:|||:||||:|||||||||||||||||||||
Db    242 LSLSTLATPKRSGNN----NGSEFSDIIPLR---TADSVFCPHYEKVSGDYGHPVYIVQ 294

Qy    327 DGPPQSPPNIY 338
          |||||||||||
Db    295 EMPPQSPANIYY 306
```

```
Db     295 EMPPQSPANIYY 306
```

```
RESULT 23
AAY71438

ID   AAY71438 standard; peptide; 89 AA.
XX
AC   AAY71438;
XX
DT   04-OCT-2000  (first entry)
XX
DE   Human ephrin B3 C-terminal cytoplasmic domain.
XX
OS   Homo sapiens.
XX
KW   PDZ domain; B class ephrin; Eph receptor tyrosine kinase; RTK; modulator;
KW   cellular process; nerve cell interaction; regeneration of nerve cell;
KW   axonogenesis; antiproliferative; proliferative disorder; treatment;
KW   differentiative disorder; human; cytoplasmic domain.
XX
PN   WO200031124-A2.
XX
PD   02-JUN-2000.
XX
PF   19-NOV-1999;  99WO-CA01101.
XX
PR   20-NOV-1998;  98US-0109158.
XX
PA   (MOUN ) MOUNT SINAI HOSPITAL.
XX
PI   Lin D, Pawson A;
XX
DR   WPI; 2000-400038/34.
XX
PT   Isolated complex for treating proliferative or differentiative
PT   disorders comprises B class ephrin and PDZ domain containing protein -
PT   disorders comprises B class ephrin and PDZ domain containing protein.
XX
PS   Example; Fig 1; 59pp; English.
XX
CC   The patent discloses a complex comprising of a B class ephrin and PDZ
CC   domain containing protein. B class ephrins function as ligands for Eph
CC   receptor tyrosine kinases (RTK) and possess a transmembrane element and a
CC   highly conserved cytoplasmic tail at the C-terminus, that contains a PDZ
CC   binding site. This complex is used in methods to modulate the interaction
CC   of a B class ephrin and PDZ domain containing protein and to identify
CC   modulator compounds. It is also used for modulating cellular processes
CC   like, axonogenesis, nerve cell interactions and regeneration of nerve
CC   cells. The complex is also useful for treating proliferative or
CC   differentiative disorders associated with this protein complex.
CC   The present sequence is the human ephrin B3 C-terminal cytoplasmic
CC   domain, comprising conserved sequences reminiscent of known or predicted
CC   binding sites for PDZ domains. Ephrin B3 is also known as NLERK-2,
```

```
FH   Key              Location/Qualifiers
FT   Region           57..75
FT                    /note= "Conserved region"
FT   Modified-site    60
FT                    /note= "Conserved tyrosine phosphorylation site"
FT   Modified-site    67
FT                    /note= "Conserved tyrosine phosphorylation site"
FT   Modified-site    72
FT                    /note= "Conserved tyrosine phosphorylation site"
FT   Region           78..82
FT                    /note= "Conserved region"
FT   Region           84..89
FT                    /note= "Conserved region"
FT   Domain           86..89
FT                    /label= PDZ_domain
FT   Modified-site    86
FT                    /note= "Conserved tyrosine phosphorylation site"
FT   Modified-site    87
FT                    /note= "Conserved tyrosine phosphorylation site"
```

```
CC  Elk-L3, EFL-6, ELF-3 and LERK-8.
XX
SQ  Sequence  89 AA;

Query Match
Best Local Similarity  97.7%;  Score 483; DB 21; Length 89;
Matches  85; Conservative  1; Mismatches 1; Indels 0; Gaps 0;

Qy  252 RRRRAKPSESRHPGPGSFGRGGSLGLGGGGGMGPREAAPEGELGIALRGGGAADPPFCPHY 311
         |||||||||||||||||||||||||||||||||||||||||||||||||||||  |||||||
Db    1 RRRRAKPSESRHPGPGSFGRGGSLGLGGGGGMGPREAAPEGELGIALRGGGTADPPFCPHY 60

Qy  312 EKVSGDYGHPVYIVQDGPPQSPPNIIY 338
         ||||||||||||||||||||||||||
Db   61 EKVSGDYGHPVYIVQDGPPQSPPNIIY 87

RESULT 24
AAY96782
ID  AAY96782 standard; Protein; 658 AA.
XX
AC  AAY96782;
XX
DT  26-SEP-2000  (first entry)
XX
DE  Ephrin-B2-Ephrin-B2-FC fusion protein.
XX
KW  Ephrin-B2; ELK receptor; ligand; dimer; Fc domain; fusion protein;
KW  Efl-6 antagonist; neurological.
XX
OS  Chimeric - Homo sapiens.
OS  Chimeric - Synthetic.
XX
FH  Key  Location/Qualifiers
FH
FT  Peptide    1..30
FT            /label= Signal_peptide
FT            /note="Derived from ephrin-B2 ectodomain 1"
FT  Protein   31..225
FT            /label= Ephrin-B2_Ectodomain-1
FT  Peptide   226..228
FT            /label= Bridging_peptide
FT  Protein   229..423
FT            /label= Ephrin-B2_ectodomain_2
FT  Peptide   424..426
FT            /label= Bridging_peptide
FT  Region    427..658
FT            /label= Human_IgG1_Fc_region

PN  WO200037642-A1.
XX
PD  29-JUN-2000.
XX
PF  23-DEC-1999;  99WO-US30900.
XX
PR  23-DEC-1998;  98US-0113387.
XX
PA  (REGE-)  REGENERON PHARM INC.
XX
PI  Davis SJ, Gale NW, Yancopoulos GD, Stahl N;
XX
DR  WPI; 2000-442670/38.
DR  N-PSDB; AAA51346.
XX
PT  Polynucleotide encoding a fusion polypeptide, useful for promoting
PT  differential function and influencing phenotype, comprises two subunits
PT  containing at least one copy of the receptor binding domain of a ligand
XX
PS  Example 12; Fig 15A-B; 97pp; English.
XX
CC  Production of homogenous forms of clustered ligands is broadly applicable
CC  to improve the affinity and/or increase the activity of a ligand as
CC  compared to the native form of the ligand. Ephrin fusion proteins have
CC  been constructed, which may be useful for treating neurological
```

```
CC  disorders. The ephrin fusion proteins are preferably capable of binding
CC  to Elk receptor and are especially Efl-6 antagonists. The fusion proteins
CC  were constructed after it was demonstrated that similar improved
CC  activities could be achieved using Tie-2 receptor ligands.
CC  Angiopoietin-1 (Ang-1) is a Tie-2 receptor ligand and is an agonist for
CC  Tie-2, whereas angiopoietin-2 (Ang-2) is a naturally occuring antagonist
CC  of the Tie-2 receptor. The fibrinogen domains (FD) of Ang-1 and Ang-2 are
CC  the receptor-binding domain fused to an Fc domain), can bind to the Tie-2
CC  (Ang-1 fibrinogen domain fused to an Fc domain), can bind to the Tie-2
CC  receptor with much higher affinity than monomeric Ang-1-FD (dimerization
CC  occurs between the Fc components of adjacent molecules). However,
CC  Ang-1-FD-Fc is not able to induce phosphorylation (activate) the Tie-2
CC  receptor on endothelial cells unless it is further clustered with goat
CC  anti-human Fc antibodies. The novel fusion proteins, mutant versions of
CC  Ang-1-FD and Ang-2-FD, were designed that were intrinsically more highly
CC  clustered. Tie-2 agonist fusion proteins may be used as haematopoietic
CC  factors. Tie-2 receptor antagonist fusion proteins may be used to
CC  diagnose or treat, e.g. myeloproliferative or other proliferative
CC  disorders of blood forming organs, e.g. thrombocythemias, polycythemias
CC  and leukemias.
XX
SQ  Sequence  658 AA;

Query Match
Best Local Similarity  18.7%;  Score 458.5; DB 21; Length 658;
Matches  128; Conservative  40; Mismatches 111; Indels 143; Gaps 12;

Qy    5 HSGPGGVRVGALLLLGVLLGLVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDLLCPR 64
         |||  |     ||:|:|| ||||| ||| ||:| |  ::| |||| |||||||||:||:|||
Db  218 HSGNN----------LLGGPGIV---LEPIYWNSSNSKFLPGGQLVIYPQIGDKLDIICPK 265

Qy   65 ARPPGPHSSPNYEFYKLIYLVGGAQGRRCEAPPAPNLLLTCDRPDLDLRFTIKFQEYSPNL 124
         |  |     |||||:||: : |:||| || |||||||| :||||||:|| ||||||||:|||
Db  266 V---DSKIVGQYEYYKYYVMVDKDQADRCTIKKENTPLLNCARPDQDVKFTIKFQERSPNL 322

Qy  125 WGHEFRSHHDYYIIATSDGTREGLESLLQGGVCLTRGMKVLLRVGQSPRGGAVPRKPVSEM 184
         ||:||: : |||||:|: | :||| : :| |||||| |:|||:|| |
Db  323 WGLEFQKNKDYYIISTNGSLEGLDNQEGGVCQTRAMKILMKVGQD-------------- 368

Qy  185 PMERDRGAAHSLEPGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAGGLALLLLLGVAG 244
                   :| |  :| ||: ||: |            ||
Db  369 ----------ASSAGSARNHGPTRRPELB-------------------------AG 389

Qy  245 AGGAMCWRRRAKPSESRHPGPGSFGRGGSLGLGGGGGMGPREAEPGELGIALRGGGAAD 304
            ||        ||           |         :
Db  390 TNG------RSSTTSPFVKPNFGSSTDGNSAGHSGNNLL--------------GGPGE 427

Qy  305 P--------------PPCPHYEKVSG-----------------DYGH------- 320
         |              |:  : :  |                   |||
Db  428 PKSCDKHITCPPCPAPELLGGPSVFLPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFN 487

Qy  321 ------PVYIVQDGPPQSPPNIIYTSISVLEWPILIHTIQLFFMRSKCCSRVTTFLFPVQVI 374
              | :|: |   | :    :                |: |:  |  :  :  :
Db  488 WYVDGVEVHNAKTKPREEQYNSTYRVVSVL--TVLHQDWLNGKEYKC-KVSNKALPAPIE 544

Qy  375 TT 376
         ||
Db  545 KT 546

RESULT 25
AAY96781
ID  AAY96781 standard; Protein; 683 AA.
XX
AC  AAY96781;
XX
DT  26-SEP-2000  (first entry)
XX
DE  Ephrin-B1-Ephrin-B1-FC fusion protein.
XX
KW  Ephrin-B1; ELK receptor; ligand; dimer; Fc domain; fusion protein;
KW  Efl-6 antagonist; neurological.
XX
```

```
OS   Chimeric - Homo sapiens.
OS   Chimeric - Synthetic.
XX
FH   Key             Location/Qualifiers
FT
FT   Peptide         1..29
FT                   /label= Signal_peptide
FT   Peptide         30..237
FT                   /note= "derived from ELK-L Ectodomain 1"
FT   Protein         1..237
FT                   /label= ELK-L_ectodomain_1
FT   Peptide         238..240
FT                   /label= ELK-L_ectodomain_1
FT   Peptide         241..448
FT                   /label= Bridging_peptide
FT   Protein         241..448
FT                   /label= ELK-L_Ectodomain_2
FT   Peptide         449..451
FT                   /label= ELK-L_Ectodomain_2
FT   Peptide         452..683
FT                   /label= Bridging_peptide
FT   Region          452..683
FT                   /label= Human_IgG1_Fc_region
XX
PA   (REGE-) REGENERON PHARM INC.
PN   Davis SJ, Gale NW, Yancopoulos GD, Stahl N;
XX
PR   23-DEC-1998; 98US-0113387.
PF   23-DEC-1999; 99WO-US30900.
PD   29-JUN-2000.
PN   WO200037642-A1.
XX
DR   WPI; 2000-442670/38.
DR   N-PSDB; AAA51345.
XX
PS   Example 12; Fig 14A-E; 97pp; English.
XX
CC   Polynucleotide encoding a fusion polypeptide, useful for promoting
CC   differential function and influencing phenotype, comprises two subunits
CC   containing at least one copy of the receptor binding domain of a ligand
XX
CC   Production of homogenous forms of clustered ligands is broadly applicable
CC   to improve the affinity and/or increase the activity of a ligand as
CC   compared to the native form of the ligand. Ephrin fusion proteins have
CC   been constructed, which may be useful for treating neurological
CC   disorders. The ephrin fusion proteins are preferably capable of binding
CC   to Elk receptor and are especially Efl-6 antagonists. The fusion proteins
CC   were constructed after it was demonstrated that similar improved
CC   activities could be achieved using Tie-2 receptor ligands.
CC   Angiopoietin-1 (Ang-1) is a Tie-2 receptor ligand and is an agonist for
CC   Tie-2, whereas angiopoietin-2 (Ang-2) is a naturally occuring antagonist
CC   for the Tie-2 receptor. The fibrinogen domains (FD) of Ang-1 and Ang-2 are
CC   of the Tie-2 receptor. The fibrinogen domains, of e.g. Ang-1-FD-Fc
CC   (Ang-1 fibrinogen domain fused to an Fc domain), can bind to the Tie-2
CC   receptor with much higher affinity than monomeric Ang-1-FD (dimerization
CC   occurs between the Fc components of adjacent molecules). However, the
CC   Ang-1-FD-Fc is not able to induce phosphorylation (activate) the Tie-2
CC   receptor on endothelial cells unless it is further clustered with goat
CC   anti-human Fc antibodies. The novel fusion proteins, mutant versions of
CC   Ang-1-FD and Ang-2-FD, were designed that were intrinsically more highly
CC   clustered. Tie-2 agonist fusion proteins may be used as haematopoietic
CC   factors. Tie-2 receptor antagonist fusion proteins or other proliferative
CC   diagnose or treat, e.g. myeloproliferative or other proliferative
CC   disorders of blood forming organs, e.g. thrombocythemias, polycythemias
CC   and leukemias.
XX
SQ   Sequence 683 AA;

Query Match      18.5%;  Score 454;  DB 21;  Length 683;
Best Local Similarity    44.0%;  Pred. No. 1.3e-29;
Matches    92;  Conservative    32;  Mismatches    72;  Indels    13;  Gaps    3;

Qy    27  GLSLEPVYWNSANKRFQAEGGYVYIPQIGDRLDLLCPRARPPGPHSSENYEFYKLYLVGG  86
              |:|||||:|::| |:|: :| |||:||||||:|:|||||||||||
Db   240  GKNLEPVSWSSLNPKFLSGKGLIVYPKIGDKLDIICPRAEAGRP----YEYYKLYLVRP  294
```

```
Qy    87  AQGRRCEAPPAPNLLLTCDRPDLDLRFTIKFQEYSPNLWGHEPFRSHHDYYIIATSDGTRE  146
              ||:||:|||:|:|||:::|:|:
Db   295  EQAAACSTVLDPNVLVTCNRPEQEIRFTIKFQEFSPNYMGLEFKKHDYYITSRPSNGSLE  354

Qy   147  GLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKPVSEMPRDRGAAHSLE-PGKENLPG  205
              |||||:||||:|::|||||:|:|:|||:
Db   355  GLENREGGVCVRTRTMKIIMKVGQDPNAVTPEQLITSRPSKEADNTVKMATQAPGSRGGSLG  414

Qy   206  DPTSNATSRGAEGPLPPPSMPAVAGAAGG  234
              :|||| |:||
Db   415  DSDGKHETVNQEEKSGP------GASGG  436


RESULT 26
AAE24019

ID   AAE24019 standard; Protein; 229 AA.
XX
AC   AAE24019;
XX
DT   23-SEP-2002  (first entry)
XX
DE   Murine ephrin B2 ligand.
XX
KW   Murine; extracellular region; Eph B receptor; ephrin B ligand; tumour;
KW   ocular neovascularisation; gene therapy; cytostatic.
OS   Mus musculus.
XX
PN   WO200226827-A1.
XX
PD   04-APR-2002.
XX
PF   28-SEP-2001; 2001WO-EP11252.
PR   29-SEP-2000; 2000CH-0001910.
XX
DR   WPI; 2002-405047/43.
DR   N-PSDB; AAD38768.
XX
PA   (NOVS ) NOVARTIS AG.
PA   (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
PN   Martiny-baron G, Wood JM, Liau G;
XX
PS   Claim 6; Page 48-49; 55pp; English.
XX
CC   New soluble polypeptides of the extracellular region of the Eph B
CC   receptor and ephrin B ligands, useful in gene therapy for treating a
CC   tumour disease or ocular neovascularisation in mammals, including humans
XX
CC   The invention relates to new isolated soluble polypeptides comprising an
CC   amino acid sequence of an extracellular region of an Eph B receptor or an
CC   amino acid sequence of an ephrin B ligand, or essentially similar amino
CC   acid sequences of extracellular region of an Eph B receptor or an amino
CC   acid sequence of an ephrin B ligand. Extracellular region of an Eph B
CC   receptor is capable of binding with high affinity to a mammalian ephrin B
CC   ligand. Ephrin B ligand is capable of binding with high affinity to a
CC   mammalian Eph B receptor. The nucleic acid encoding the polypeptides or
CC   the polypeptides are useful for producing pharmaceutical compositions for
CC   treating a tumour disease or ocular neovascularisation in mammals,
CC   including humans. The vector is useful in producing a pharmaceutical
CC   composition for treating the diseases cited above in mammals,
CC   particularly humans, by means of gene therapy. The present sequence is
CC   murine ephrin B2 ligand.
XX
SQ   Sequence 229 AA;

Query Match      18.2%;  Score 447;  DB 23;  Length 229;
Best Local Similarity    36.4%;  Pred. No. 1.2e-29;
Matches   100;  Conservative    33;  Mismatches    80;  Indels    62;  Gaps    4;
```

```
RESULT 27
AAE24020

Qy     14  GALILLLGVLGLVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDLLCPRARPPGPHSS   73
               |:|:|       ||||:
Db     17  GLIMVLCRTAISRIVLEPIYWNSSNSKFLPGQGLVLYPQIGDKLDIICPKV---DSKTV    73

Qy     74  PNYEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDLDLRFTIKFQEYSPNLWGHEFRSHH  133
               |||:|:||||:     ||||||||||||||||
Db     74  GQYEYYKVYMVDKDQADRCTIKKENTPLLNCARPDQDVKFTIKFQEFSPNLGLEFQKNK  133

Qy    134  DYYIIATSDGTRGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKPVSEMPMERDRGAA  193
Db    134  DYYIISTSNGSLEGLDNQEGGVCCTRAMKILMKVGQD------------------      170

Qy    194  HSLEPGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAGGLALILLGVAGAGGAMCWRR  253
Db    171  -------------------------ASSAGSARNHGPTRRPELE----------AGTNG-----  194

Qy    254  RRAKPSESRHPGPGSFGRGGSLGLGGGGMGPPREA  288
               :|:|
Db    195  RSSTTSPFVKPNPGSSTDGNSAGHSGNNLLGSEVA  229


ID   AAE24020 standard; protein; 229 AA.
XX
AC   AAE24020;
XX
DT   23-SEP-2002 (first entry)
XX
DE   Human ephrin B2 ligand.
XX
KW   Human; extracellular region; Eph B receptor; ephrin B ligand; tumour;
KW   ocular neovascularisation; gene therapy; cytostatic.
XX
OS   Homo sapiens.
XX
PN   WO200226827-A1.
XX
PD   04-APR-2002.
XX
PF   28-SEP-2001; 2001WO-EP11252.
XX
PR   29-SEP-2000; 2000CH-0001910.
XX
PI   Martiny-baron G, Wood JM, Liau G;
XX
PA   (NOVS ) NOVARTIS AG.
PA   (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
DR   WPI; 2002-405047/43.
DR   N-PSDB; AAD38769.
XX
PT   New soluble polypeptides of the extracellular region of the Eph B
PT   receptor and ephrin B ligands, useful in gene therapy for treating a
PT   tumour disease or ocular neovascularization in mammals, including humans
PT   -
XX
CC   The invention relates to new isolated soluble polypeptides comprising an
CC   amino acid sequence of an extracellular region of an Eph B receptor or an
CC   amino acid sequence of an ephrin B ligand, or essentially similar amino
CC   acid sequences of extracellular region of an Eph B receptor or an amino
CC   acid sequence of an ephrin B ligand. Extracellular region of an Eph B
CC   receptor is capable of binding with high affinity to a mammalian ephrin B
CC   ligand. Ephrin B ligand is capable of binding with high affinity to a
CC   mammalian Eph B receptor. The nucleic acid encoding the polypeptides or
CC   the polypeptides are useful for producing pharmaceutical compositions for
CC   treating a tumour disease or ocular neovascularisation in mammals,
CC   including humans. The vector is useful in producing a pharmaceutical
CC   composition for treating the diseases cited above in mammals,
CC   particularly humans, by means of gene therapy. The present sequence is
CC   human ephrin B2 ligand.
XX
PS   Claim 6; Page 51-52; 55pp; English.
XX
SQ   Sequence   229 AA;
```

```
RESULT 28
AAW06333

Query Match             18.1%;  Score 443;  DB 23;  Length 229;
Best Local Similarity   35.6%;  Pred. No. 2.7e-29;
Matches    98;  Conservative   36;  Mismatches   79;  Indels   62;  Gaps    4;

Qy     14  GALILLLGVLGLVSGLSLEPVVWNSANKRFQAEGGYVLYPQIGDRLDLLCPRARPPGPHSS   73
               |:|:|     |||||:
Db     14  GVLMVLCRTAISKSIVLEPIYWNSSNSKFLPGQGLVLYPQIGDKLDIICPKV---DSKTV   70

Qy     74  PNYEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDLDLRFTIKFQEYSPNLWGHEFRSHH  133
               |||:||||:    ||||||||||||||||
Db     71  GQYEYYKVYMVDKDQADRCTIKKENTPLLNCAKPDQDIKFTIKFQEFSPNLWGLEFQKNK  130

Qy    134  DYYIIATSDGTRGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKPVSEMPMERDRGAA  193
Db    131  DYYIISTSNGSLEGLDNQEGGVCVCTRAMKILMKVGQ------------------      166

Qy    194  HSLEPGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAGGLALILLGVAGAGGAMCWRR  253
Db    167  ------------------DASSAGSTRNKDPTRRPELE----------AGTNG-----      191

Qy    254  RRAKPSESRHPGPGSFGRGGSLGLGGGGMGPPREA  288
Db    192  RSSTTSPFVKPNPGSSTDGNSAGHSGNNLLGSEVA  226


ID   AAW06333 standard; protein; 195 AA.
XX
AC   AAW06333;
XX
DT   17-JAN-1997 (first entry)
XX
DE   Ligand #1 for receptor-type tyrosine kinase protein.
XX
KW   Receptor-type tyrosine kinase; ligand; coomassie staining; PAS staining;
KW   human.
XX
OS   Homo sapiens.
XX
PN   JP08188596-A.
XX
PD   23-JUL-1996.
XX
PF   13-JAN-1995; 95JP-0003677.
XX
PR   09-NOV-1994; 94JP-0275411.
PR   19-OCT-1994; 94JP-0253848.
XX
PA   (ASAH ) ASAHI KASEI KOGYO KK.
XX
DR   WPI; 1996-388601/39.
XX
PT   New ligand for receptor type tyrosine kinase - has mol.wt. 22-25
PT   kilo:dalton(s) and is positive for Coomassie and PAS staining
XX
CC   AAW06333 and AAW06334 represent receptor-type tyrosine kinase receptor
CC   binding ligands of the invention. These sequences both have the
CC   N-terminal sequence represented by AAW06332. The ligands of the
CC   invention recognise the fragment represented by AAW06330 (see AAW06330
CC   represented by AAW06330 (see AAW06331 for full length sequence). The
CC   proteins of the invention have a molecular weight of 23500 (plus or minus
CC   1500) Da, and are positive for Coomassie staining and PAS staining. The
CC   protein is a new ligand of receptor-type tyrosine kinase, and can be
CC   prepared by standard recombinant techniques.
XX
PS   Claim 5; Page 43; 51pp; Japanese.
XX
SQ   Sequence   195 AA;
```

Query Match   17.6%;   Score 431.5;   DB 17;   Length 195;
Best Local Similarity   45.3%;   Pred. No. 2e-28;
Matches   87;   Conservative   32;   Mismatches   66;   Indels   7;   Gaps   4;

```
Qy     30 LEPVYWNSANRFQAEGGYVLYPQIGDRLDLLCPRARPPGPHSSSPNYEFYKLIYLVGGAQG 89
          ||||:|||||||:||||:|:||||||:||:|:|||||:||||||||:||||||||||||
Db      5 LEPIYWNSNSKFLPCGGLVLYPQIGDKLDIICPKV---DSKTVGGYEYYKVYMVDKDQA 61

Qy     90 RRCEAPPAPNLLITCDRPDLDLRFTIKFQEYSPNIWGHEFRSHHDYYIIATSDGTREGLE 149
          ||:|||||| | |||||:|:|:|||:::|||:| | ||::|||||||||:|:||:|||:
Db     62 DRCTIKKENTPLLNCAKPDQODIKFTIKFQEFSPNLWGLEFQKNKDYYIISTSNGSLEGLD 121

Qy    150 SLQGGVCLTRGMKVLLRVGQ--SPRGGAVPRKPVSEMPMER-DRGAAHSLEPGKENLPGD 206
          |||||||| ||||||||:|||  :|||||||:||:||:|:|:  :|:||:||||||||:|:
Db    122 NQEGGVCQTPAMKILMKVGQDASSAGSTRNKDPTRRPELEAGTNGRSSTTSPFVKPNPGS 181

Qy    207 PTSNATSRGAEG 218
          |||||||||||
Db    182 ST-DGNSAGHSG 192
```

RESULT 29
AAW11307
ID   AAW11307 standard; protein; 195 AA.
XX
AC   AAW11307;
XX
DT   21-MAR-1997   (first entry)
XX
DE   Receptor-type tyrosine kinase #3.
XX
KW   Receptor-type tyrosine kinase; ligand; autophosphorylation;
KW   insect cell; animal cell; growth promoter; undifferentiated cell;
KW   insulin; interleukin; fibroblast growth factor; hepatocyte growth factor;
KW   nerve growth factor; interferon-gamma; tumour necrosis factor; inducer.
XX
OS   Homo sapiens.
XX
PN   JP08280385-A.
XX
PD   29-OCT-1996.
XX
PF   13-APR-1995; 95JP-0087878.
PR   13-APR-1995; 95JP-0087878.
XX
PA   (ASAH) ASAHI KASEI KOGYO KK.
XX
DR   WPI; 1997-014848/02.
XX
PT   LPM cell culture medium - for culturing cells which produce
PT   receptor-type tyrosine kinase ligand
XX
PS   Disclosure; Page 38-39; 45pp; Japanese.
XX
CC   AAW11307 and AAW11308 represent receptor-type tyrosine kinases. These
CC   sequences can be included in the LPM medium of the invention. The
CC   medium is used to culture cells capable of producing a ligand for
CC   receptor-type tyrosine kinase. The ligand serves as a physiologically
CC   active substance to induce enzyme activity and autophosphorylation of
CC   tyrosine kinase, which is involved in the development and differentiation
CC   of animal and insect cells. The ligand can also be used as in a
CC   pharmaceutical preparation for promoting growth of undifferentiated
CC   cells. The medium may also include an inducer, such as insulin, an
CC   interleukin, a fibroblast growth factor, or interferon-gamma.
XX
SQ   Sequence   195 AA;

Query Match   17.6%;   Score 431.5;   DB 18;   Length 195;
Best Local Similarity   45.3%;   Pred. No. 2e-28;
Matches   87;   Conservative   32;   Mismatches   66;   Indels   7;   Gaps   4;

```
Qy     30 LEPVYWNSANRFQAEGGYVLYPQIGDRLDLLCPRARPPGPHSSSPNYEFYKLIYLVGGAQG 89
```

RESULT 30
AAM37671
ID   AAM37671 standard; Protein; 92 AA.
XX
AC   AAM37671;
XX
DT   17-OCT-2001   (first entry)
XX
DE   Peptide #11708 encoded by probe for measuring placental gene expression.
XX
KW   Probe; microarray; human; placenta; antenatal diagnosis;
KW   genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200157272-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00663.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,   Hanzel DK,   Chen W,   Rank DR;
XX
DR   WPI; 2001-488897/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human placenta -
XX
PS   Claim 27; SEQ ID NO 37940; 654pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes (SENP:
CC   see AAI31315-AAI57546). The present sequence is a peptide encoded by one
CC   such probe. The probes are useful for producing a microarray for
CC   predicting, measuring and displaying gene expression in samples derived
CC   from human placenta. The probes are useful for antenatal diagnosis of
CC   human genetic disorders.
XX
SQ   Sequence   92 AA;

Query Match   11.6%;   Score 284.5;   DB 22;   Length 92;
Best Local Similarity   54.2%;   Pred. No. 2e-16;
Matches   52;   Conservative   15;   Mismatches   24;   Indels   5;   Gaps   1;

```
Qy     42 FQAEGGYVLYPQIGDRLDLLCPRARPPGPHSSSPNYEFYKLIYLVGGAQGRRCEAPPAPNLL 101
          :  | | ||||||:|:||||||||::|||||  ||||||||||  | :| | ||||||:|:|
Db      1 FLSGKGLVIYPKIGDRLDIICPRAEAGRP----YEYYKLYLVRPEQAAACSTVLDPNVL 55
```

Second column bottom (RESULT 29 match detail):

```
Db      5 LEPIYWNSNSKFLPCGGLVLYPQIGDKLDIICPKV---DSKTVGGYEYYKVYMVDKDQA 61

Qy     90 RRCEAPPAPNLLITCDRPDLDLRFTIKFQEYSPNIWGHEFRSHHDYYIIATSDGTREGLE 149
Db     62 DRCTIKKENTPLLNCAKPDQODIKFTIKFQEFSPNLWGLEFQKNKDYYIISTSNGSLEGLD 121

Qy    150 SLQGGVCLTRGMKVLLRVGQ--SPRGGAVPRKPVSEMPMER-DRGAAHSLEPGKENLPGD 206
Db    122 NQEGGVCQTPAMKILMKVGQDASSAGSTRNKDPTRRPELEAGTNGRSSTTSPFVKPNPGS 181

Qy    207 PTSNATSRGAEG 218
Db    182 ST-DGNSAGHSG 192

Qy     30 LEPVYWNSANRFQAEGGYVLYPQIGDRLDLLCPRARPPGPHSSSPNYEFYKLIYLVGGAQG 89
```

```
Qy    102 LTCDRPDLDLRFTIKFQEYSPNLWGHEFRSHHDYYI 137
          :|:||:|||||:|||:|:|||  ||:||:||||||||
Db     56 VTCNRPEQEIRFTIKFQEFSPNYMGLEFKXHHDYYI 91
```

RESULT 31
ABG46524
ID   ABG46524 standard; Peptide; 92 AA.
AC   ABG46524;
XX
DT   19-AUG-2002 (first entry)
XX
DE   Human peptide encoded by genome-derived single exon probe SEQ ID 36189.
XX
KW   Human; single exon probe; asthma; lung cancer; COPD; ILD;
KW   chronic obstructive pulmonary disease; interstitial lung disease;
KW   familial idiopathic pulmonary fibrosis; neurofibromatosis;
KW   tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
KW   Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
KW   pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
KW   pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KW   primary ciliary dyskinesis; pulmonary hypertension;
KW   hyaline membrane disease.
OS   Homo sapiens.
XX
PN   WO200186003-A2.
XX
PD   15-NOV-2001.
XX
PF   30-JAN-2001; 2001WO-US00665.
XX
PR   04-FEB-2000; 2000US-180312P.
PR   26-MAY-2000; 2000US-207456P.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-234687P.
PR   27-SEP-2000; 2000US-236359P.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR   WPI; 2002-114183/15.
XX
PT   Spatially-addressable set of single exon nucleic acid probes, used to
PT   measure gene expression in human lung samples -
XX
PS   Claim 27; SEQ ID NO 36189; 634pp; English.
XX

CC   The invention relates to a spatially-addressable set of single exon
CC   nucleic acid probes for measuring gene expression in a sample derived
CC   from human lung comprising single exon nucleic acid probes having one of
CC   12614 nucleic acid sequences mentioned in the specification, or their
CC   complements or the 12387 open reading frames derived from the 12614
CC   probes. Also included are a microarray comprising the novel set of
CC   probes; the novel set of probes which hybridise at high stringency to a
CC   nucleic acid expressed in the human lung; measuring gene expression in a
CC   sample derived from human lung, comprising (a) contacting the array with
CC   a collection of detectably labeled nucleic acids derived from human lung
CC   mRNA, and (b) measuring the label detectably bound to each probe of
CC   the array; identifying exons in a eukaryotic genome, comprising
CC   (a) algorithmically predicting at least one exon from genomic sequences
CC   of the eukaryote; and (b) detecting specific hybridisation of detectably
CC   labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
CC   having a fragment identical to the predicted exon, the probe is included
CC   in the above mentioned microarray; assigning exons to a single gene,
CC   comprising (a) identifying exons from genomic sequence by the method
CC   above and (b) measuring the expression of each of the exons in several
CC   tissues and/or cell types using hybridisation to a single exon
CC   microarrays having a probe with the exon, where a common pattern of
CC   expression of the exons in the tissues and/or cell types indicates that
CC   the exons should be assigned to a single gene; a peptide comprising one
CC   of 12001 sequences, mentioned in the specification, or encoded by the
CC   probes/open reading frames (ORF). The probes are used for gene
CC   expression analysis, and for identifying exons in a gene, particularly
CC   using human derived mRNA and for the study of lung diseases
CC   such as asthma, lung cancer, chronic obstructive pulmonary disease
CC   (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
CC   fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
CC   Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
CC   haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,
CC   pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic
CC   pulmonary dysplasia, primary ciliary dyskinesis, pulmonary hypertension
CC   and hyaline membrane disease. The present sequence is a peptide/protein
CC   encoded by a single exon probe of the invention.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 92 AA;

Query Match   11.6%; Score 284.5; DB 23; Length 92;
Best Local Similarity   54.2%; Pred. No. 2e-16;
Matches   52; Conservative 15; Mismatches 24; Indels   5; Gaps   1;

Qy   42 FQAEGGYVLYPQIGDRLDLLCPRARPPGPHSSPNYEFYKLYLVGGAQGRRCEAPAPNLL 101
         |:| ||||||||:|:||| |||||| ||||| |||||||||||||| ||:|
Db   1 FLSGKGLVIYPXIGDKLDIICPRAEAGRP-----YEYYKLYLVRPEQAAACSTVLDPNVL 55

Qy   102 LTCDRPDLDLRFTIKFQEYSPNLWGHEFRSHHDYYI 137
          :|:||:|||||:|||:|:||| ||:||:||||||||
Db   56 VTCNRPEQEIRFTIKFQEFSPNYMGLEFKXHHDYYI 91

RESULT 32
AAM37534
ID   AAM37534 standard; Protein; 136 AA.
AC   AAM37534;
XX
DT   17-OCT-2001 (first entry)
XX
DE   Peptide #11571 encoded by probe for measuring placental gene expression.
XX
KW   Probe; microarray; human; placenta; antenatal diagnosis;
KW   genetic disorder.
OS   Homo sapiens.
XX
PN   WO200157272-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00663.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR   WPI; 2001-488897/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human placenta -
XX
```

**Left column**

```
PS   Claim 27; SEQ ID No 37803; 654pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes (SENP:
CC   see AAI3315-AAI57546). The present sequence is a peptide encoded by one
CC   such probe. The probes are useful for producing a microarray for
CC   predicting, measuring and displaying gene expression in samples derived
CC   from human placenta. The probes are useful for antenatal diagnosis of
CC   human genetic disorders.
XX
SQ   Sequence  136 AA;

Query Match          8.0%;  Score 196.5; DB 22; Length 136;
Best Local Similarity 36.8%;  Pred. No. 8.8e-09;
Matches 50; Conservative 13; Mismatches 34; Indels 39; Gaps 5;

Qy  230 GAAGGLA------------------LLILGVAGAGGA---------MCWRRRRAKPSESRHPGP 266
        |:|| :                   ||||                 :| :
Db   11 GASGGSSGDPDGFFNSKVALFAAVGAGCVIFLLIIFLTVLLLKLRKRHRKHTQQ----- 65

Qy  267 GSFGRGGSLGL----GGGGGMGPREAEPGELGIALRGGGAADPPFCPHYEKVSGDYGHPV 322
        :|| | ||      |||||||||||||||||
Db   66 -----RAAALSLSTLASPKGGSGGTAGTEPSDIIIPLR---TTENNYCPHYEKVSGDYGHPV 118

Qy  323 YIVQDGPPQSPPNIYY 338
        |||||||||| ||||
Db  119 YIVQEMPQSPANIYY 134
```

```
RESULT 33
ABG46394
ID   ABG46394 standard; Peptide; 136 AA.
XX
AC   ABG46394;
XX
DT   19-AUG-2002 (first entry)
XX
DE   Human peptide encoded by genome-derived single exon probe SEQ ID 36059.
XX
KW   Human; single exon probe; asthma; lung cancer; COPD; ILD;
KW   chronic obstructive pulmonary disease; interstitial lung disease;
KW   familial idiopathic pulmonary fibrosis; neurofibromatosis;
KW   tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
KW   Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
KW   pulmonary histiocytosis; lymphangioleiomyomctosis; Karagener syndrome;
KW   pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KW   primary ciliary dyskinesis; pulmonary hypertension;
KW   hyaline membrane disease.
XX
OS   Homo sapiens.
XX
PN   WO200186003-A2.
XX
PD   15-NOV-2001.
XX
PF   30-JAN-2001; 2001WO-US00665.
XX
PR   04-FEB-2000; 2000US-180312P.
PR   26-MAY-2000; 2000US-207456P.
PR   30-JUN-2000; 2000US-0608808.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-234887P.
PR   27-SEP-2000; 2000US-236359P.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
PI   Penn SG, Hanzel DK, Chen W, Rank DR;
XX
PT   Spatially-addressable set of single exon nucleic acid probes, used to
PT   measure gene expression in human lung samples -
XX
DR   WPI; 2002-114183/15.
XX
```

Right column

```
PS   Claim 27; SEQ ID No 36059; 634pp; English.
XX
CC   The invention relates to a spatially-addressable set of single exon
CC   nucleic acid probes for measuring gene expression in a sample derived
CC   from human lung comprising single exon nucleic acid probes having one of
CC   12614 nucleic acid sequences mentioned in the specification, or their
CC   complements or the 12387 open reading frames derived from the 12614
CC   probes. Also included are a microarray comprising the novel set of
CC   probes; the novel set of probes which hybridise at high stringency to a
CC   nucleic acid expressed in the human lung; measuring gene expression in a
CC   sample derived from human lung, comprising (a) contacting the array with
CC   a collection of detectably labelled nucleic acids derived from human lung
CC   mRNA, and (b) measuring the label detectably bound to each probe of
CC   the array; identifying exons in a eukaryotic genome, comprising
CC   (a) algorithmically predicting at least one exon from genomic sequences
CC   of the eukaryote; and (b) detecting specific hybridisation of detectably
CC   labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
CC   having a fragment identical to the predicted exon, the probe is included
CC   in the above mentioned microarray; assigning exons to a single gene,
CC   comprising (a) identifying exons from genomic sequence by the method
CC   above and (b) measuring the expression of each of the exons in several
CC   tissues and/or cell types using hybridisation to a single exon
CC   microarrays having a probe with the exon, where a common pattern of
CC   expression of the exons in the tissues and/or cell types indicates that
CC   the exons should be assigned to a single gene; a peptide comprising one
CC   of 12011 sequences, mentioned in the specification, or encoded by the
CC   probes/open reading frames (ORF). The probes are used for gene
CC   expression analysis, and for identifying exons in a gene, particularly
CC   using human lung derived mRNA and for the study of lung diseases
CC   such as asthma, lung cancer, chronic obstructive pulmonary disease
CC   (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
CC   fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
CC   Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
CC   haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomtosis,
CC   pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic
CC   pulmonary dysplasia, primary ciliary dyskinesis, pulmonary hypertension
CC   and hyaline membrane disease. The present sequence is a peptide/protein
CC   encoded by a single exon probe of the invention.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence  136 AA;

Query Match          8.0%;  Score 196.5; DB 23; Length 136;
Best Local Similarity 36.8%;  Pred. No. 8.8e-09;
Matches 50; Conservative 13; Mismatches 34; Indels 39; Gaps 5;

Qy  230 GAAGGLA------------------LLILGVAGAGGA---------MCWRRRRAKPSESRHPGP 266
        |:|| :                   ||||                 :| :
Db   11 GASGGSSGDPDGFFNSKVALFAAVGAGCVIFLLIIFLTVLLLKLRKRHRKHTQQ----- 65

Qy  267 GSFGRGGSLGL----GGGGGMGPREAEPGELGIALRGGGAADPPFCPHYEKVSGDYGHPV 322
        :|| | ||      |||||||||||||||||
Db   66 -----RAAALSLSTLASPKGGSGGTAGTEPSDIIIPLR---TTENNYCPHYEKVSGDYGHPV 118

Qy  323 YIVQDGPPQSPPNIYY 338
        |||||||||| ||||
Db  119 YIVQEMPQSPANIYY 134
```

```
RESULT 34
AAY71437
ID   AAY71437 standard; peptide; 82 AA.
XX
AC   AAY71437;
XX
DT   04-OCT-2000 (first entry)
XX
DE   Human ephrin B2 C-terminal cytoplasmic domain.
XX
KW   PDZ domain; B class ephrin; Eph receptor tyrosine kinase; RTK; modulator;
```

KW cellular process; nerve cell interaction; regeneration of nerve cell;
KW axonogenesis; antiproliferative; proliferative disorder; treatment;
KW differentiative disorder; human; cytoplasmic domain.
XX
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FT Region   50..68
FT   /note= "Conserved region"
FT Modified-site   53
FT   /note= "Conserved tyrosine phosphorylation site"
FT Modified-site   60
FT   /note= "Conserved tyrosine phosphorylation site"
FT Modified-site   65
FT   /note= "Conserved tyrosine phosphorylation site"
FT Region   71..75
FT   /note= "Conserved region"
FT Region   77..82
FT   /note= "Conserved region"
FT Domain   79..82
FT   /label= PDZ_domain
FT Modified-site   79
FT   /note= "Conserved tyrosine phosphorylation site"
FT Modified-site   80
FT   /note= "Conserved tyrosine phosphorylation site"
XX
PN WO20003031124-A2.
XX
PD 02-JUN-2000.
XX
PF 19-NOV-1999; 99WO-CA01101.
XX
PR 20-NOV-1998; 98US-0109158.
XX
PA (MOUN) MOUNT SINAI HOSPITAL.
PI Lin D, Pawson A;
XX
PT Isolated complex for treating proliferative or differentiative
PT disorders comprises B class ephrin and PDZ domain containing protein -
PS Example; Fig 1; 59pp; English.
XX
CC The patent discloses a complex comprising of a B class ephrin and PDZ
CC domain containing protein. B class ephrins function as ligands for Eph
CC receptor tyrosine kinases (RTK) and possess a transmembrane element and a
CC highly conserved cytoplasmic tail at the C-terminus, that contains a PDZ
CC binding site. This complex is used in methods to modulate the interaction
CC of a B class ephrin and PDZ domain containing protein and to identify
CC modulator compounds. It is also used for modulating cellular processes
CC like, axonogenesis, nerve cell interactions and regeneration of nerve
CC cells. The complex is also useful for treating proliferative or
CC differentiative disorders associated with this protein complex.
CC The present sequence is the human ephrin B2 C-terminal cytoplasmic
CC domain, comprising conserved sequences reminiscent of known or predicted
CC binding sites for PDZ domains. Ephrin B2 is also known as Htk-L, ELF-2,
CC LERK-5 and NLERK-1.
XX
DR WPI; 2000-400038/34.
XX
SQ Sequence   82 AA;

RESULT 35
AAB54187

Query Match   8.0%; Score 195; DB 21; Length 82;
Best Local Similarity   65.0%; Pred. No. 6.2e-09;
Matches   39; Conservative 4; Mismatches 11; Indels 6; Gaps 2;

Qy   279 GGGGGMGPREAEPGEELGIALRGGGAADPPFCPHYEKVSGDYGHPVYIVQDGPPQSPPNIYY 338
           ::|| ::| | :: :|: |||||||||||||||||||||||||||:|||:|||
Db   27 GGNNNG---SEPSDVIFLR---TADSVFCPHYEKVSGDYGHPVYIVQEMPPQSPANIYY 80

ID AAB54187 standard; protein; 106 AA.
XX
AC AAB54187;
XX
DT 09-MAR-2001 (first entry)
XX
DE Human pancreatic cancer antigen protein sequence SEQ ID NO:639.
XX
KW Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
KW detection; diagnosis; identification; cytostatic; neuroprotective;
KW nootropic; immunomodulatory; relaxant; contraceptive; gynaecological;
KW antiinflammatory; cardiant; gene therapy; chromosome mapping;
KW linkage analysis; tissue identification; tissue typing; forensic;
KW neural; immune system; muscular; reproductive; gastrointestinal;
KW pulmonary; cardiovascular; renal; proliferative.
XX
OS Homo sapiens.
XX
PN WO200055320-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05989.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Ruben SM;
XX
PT New nucleic acid that is a pancreatic cancer antigen for preventing,
PT treating, or ameliorating a medical condition, particular pancreatic
PT cancer, or for use in assays for diagnosing a pathological condition -
PS Claim 11; Page 1078; 1379pp; English.
XX
CC AAC98773 to AAC99231 encode the human pancreatic cancer associated
CC proteins, called pancreatic cancer antigens, given in AAB54008 to
CC AAB54466. The human pancreatic cancer antigens have cytostatic,
CC neuroprotective, nootropic, immunomodulatory, relaxant, contraceptive,
CC gynaecological, cardiant and antiinflammatory activities, and can be used
CC in gene therapy. The polynucleotide and proteins can be used for
CC preventing, treating, or ameliorating a medical condition or in assays
CC for diagnosing a pathological condition or a susceptibility to one in a
CC subject. Binding Partners to the proteins and the activity of the
CC proteins can be identified. The pancreatic cancer antigens can be used to
CC detect, treat or prevent pancreatic disorders, especially cancer.
CC Agonists and antagonists to the antigens can be screened for. The
CC pancreatic cancer antigen polynucleotides can be used to design nuclei
CC acid hybridisation probes that can be used in chromosome mapping, linkage
CC analysis, tissue identification and/or typing and a variety of forensic
CC and diagnostic methods. The proteins can be used to generate antibodies
CC which are used to purify, detect and target the polypeptides, including
CC both in vivo and in vitro diagnostic and therapeutic methods. The
CC proteins can be used to treat or prevent neural, immune system, muscular,
CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
CC sequences used in the exemplification of the present invention.
XX
SQ Sequence   106 AA;

Query Match   7.9%; Score 193.5; DB 21; Length 106;
Best Local Similarity   42.6%; Pred. No. 1.2e-08;
Matches   46; Conservative 11; Mismatches 26; Indels 25; Gaps 4;

Qy   235 LALILLLGVAGAGGAMCWRRRAKPSESRHPGPGSFGRGGSLGL----GGGGGMGPREAEP 290
           :| .|..| | | | || |: ||| |||||||||||
Db   18 LTVLLLKL-------RKRHRKHTQQ------RAAALSLSTLASPKGGSGYRAGTEP 59

Qy   291 GEELGIALRGGGAADPPFCPHYEKVSGDYGHPVYIVQDGPPQSPPNIYY 338

RESULT 36

```
Db        60 SDIIPLR--TTENNYCPHYEKVSGDYGHPVYIVQEMPPQSPANIYY 104
             ::: |||    :  |||||||||||||||||||:.|||: ||||: ||||
```

ID   AAY71436 standard; peptide; 82 AA.
XX
AC   AAY71436;
XX
DT   04-OCT-2000 (first entry)
XX
DE   Human ephrin B1 C-terminal cytoplasmic domain.
XX
KW   PDZ domain; B class ephrin; Eph receptor tyrosine kinase; RTK; modulator;
KW   cellular process; nerve cell interaction; regeneration of nerve cell;
KW   axonogenesis; antiproliferative; proliferative disorder; treatment;
KW   differentiative disorder; human; cytoplasmic domain.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FH
FT   Region   50..68
FT   /note= "Conserved region"
FT   Modified-site 53
FT   /note= "Conserved region"
FT   Modified-site 60
FT   /note= "Conserved tyrosine phosphorylation site"
FT   Modified-site 65
FT   /note= "Conserved tyrosine phosphorylation site"
FT   Region   71..75
FT   /note= "Conserved region"
FT   Region   77..82
FT   /note= "Conserved region"
FT   Region   79..82
FT   /note= "Conserved region"
FT   Domain   79..82
FT   /label= PDZ_domain
FT   /note= "PDZ domain"
FT   Modified-site 79
FT   /note= "Conserved tyrosine phosphorylation site"
FT   Modified-site 80
FT   /note= "Conserved tyrosine phosphorylation site"
XX
PN   WO20003124-A2.
XX
PD   02-JUN-2000.
XX
PF   19-NOV-1999;   99WO-CA01101.
XX
PR   20-NOV-1998;   98US-0109158.
XX
PS   Example; Fig 1; 59pp; English.
XX
PA   (MOUN) MOUNT SINAI HOSPITAL.
XX
PI   Lin D, Pawson A;
XX
DR   WPI; 2000-400038/34.
XX
PT   Isolated complex for treating proliferative or differentiative
PT   disorders comprises B class ephrin and PDZ domain containing protein -
XX
CC   The patent discloses a complex comprising of a B class ephrin and PDZ
CC   domain containing protein. B class ephrins function as ligands for Eph
CC   receptor tyrosine kinases (RTK) and possess a transmembrane element and a
CC   highly conserved cytoplasmic tail at the C-terminus, that contains a PDZ
CC   binding site. This complex is used in methods to modulate the interaction
CC   of a B class ephrin and PDZ domain containing protein and to identify
CC   modulator compounds. It is also used for modulating cellular processes
CC   like, axonogenesis, nerve cell interactions and regeneration of nerve
CC   cells. The complex is also useful for treating proliferative or
CC   differentiative disorders associated with this protein complex.
CC   The present sequence is the human ephrin B1 C-terminal cytoplasmic
CC   domain, comprising conserved sequences reminiscent of known or predicted
CC   binding sites for PDZ domains. Ephrin B1 is also known as LERK-2, Elk-L,
CC   EFL-3, Cek-L and STRA1.
XX
SQ   Sequence 82 AA;

Query Match   7.8%; Score 192; DB 21; Length 82;
Best Local Similarity 46.2%;   Pred. No. 1.1e-08;
Matches 42; Conservative 9; Mismatches 24; Indels 16; Gaps 3;

```
Qy  252 RRRRAKPSESRHPGPGSFGRGGSLGL----GGGGGMGPREAEPGELGTALRGGGAADPPF 307
        |||| |:|                    |||||| |  |||||  :  ||| |  |
Db    2 RKRHRKHTQQ----RAAALSLSTLASPKGGGSGTAGTEPSDIIIPLR---TTENNY 49

Qy  308 CPHYEKVSGDYGHPVYIVQDGPPQSPANIYY 338
        |||||||||||||||||||| ||||||||||
Db   50 CPHYEKVSGDYGHPVYIVQEMPPQSPANIYY 80
```

RESULT 37

ID   ABB58621 standard; Protein; 652 AA.
XX
AC   ABB58621;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 2655.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO20017042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001;   2001WO-US09231.
XX
PR   23-MAR-2000;   2000US-1916379P.
PR   11-JUL-2000;   2000US-0614150.
XX
PA   (PEKE) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL02724.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 2655; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL01840-ABL16175) and the encoded DNA
CC   sequences (ABL16176-ABL30511), expressed
CC   sequences (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 652 AA;

Query Match   7.8%; Score 192; DB 22; Length 652;
Best Local Similarity 21.2%;   Pred. No. 1.6e-07;
Matches 103; Conservative 64; Mismatches 154; Indels 164; Gaps 22;

```
Qy  33 VYWNSANKRFQAEG-GYVLYPQIG------DRLDLLCPRARPGPHSSPNYEPYKLYLVG 85
```

```
Db  219  MHWNTSNSIFRIDNTDHIIDVNKGNLAFEFDQVHIICP-VYEPGTFENET-EKYIIYNVS   276
         ::|||:    :    ::||
Qy   86  GAQGRRCEAPPA-PNLLLTCDRPDLDLRFTIKFQEYSPNLWGHEFRSHHDYYIIATSDGT   144
         ||||  :     |||        :|||
Db  277  KVEYETCRITNADPRVIAICDKPQKLMFFITFRRPFTPQPGGLEFLPGNDYYFISTS--S   334
              |:         ||| |
Qy  145  REGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKPVYSEMPMERDRGAAHSL-----EPG   199
                |||            ||:|
Db  335  KDDLYRRIGGRCSTNNMKVVFKVCCAPEDNN-KTTALSNSKSVTDTGGAINVNIANNDES   393
            ::          |:|||
Qy  200  KENLPGDPTSNAISRGAEG---------------PLPP----------P            223
         :|:
Db  394  HVNSHGNNIAIGTNIGINGGQIIGGPQSAGIPINPLSGNNNINGIPTTINSNIDQFNRIP   453
Qy  224  SMPAVAGAAGGLALLLLGVAGAGGAMCWRRRAKPSESRHPGPG-SFGRGGSLGLGGGGG    282
                                                    :
Db  454  IQPNIIGNHVGTNAVGTGIVGGGGIIL------------------TPGHAHGNINMLQPGRGGI   499
Qy  283  MGPREAEPG----ELGIALRG--------------------GGAADPPPCPHYEKVSGDYGH   320
         ::|                                       |
Db  500  NG------AYPGHHHIQTGIRINNVPTQHNYPSHKGNANSNINGNDDH------HHYNK------H   547
Qy  321  PVYIVQD------GPPQSPNIYYTSISVLEWPILHTIQLFFMRSKCSRVTTFLFPVQV     373
         |:|                                            |:
Db  374  ITTSTCRMTSFSFTTLNPSMQACRAQMGEFRIRWCFWGDRILGTALFVLVLILLLGRLNM   433
Qy  374  PNEVVRNEELTYNSGAATSDGNIFALWIWLS--------------------IFP--L     583
         :|:      |||:|
Db  584  LSIQSCHLSSY--------------WI-------SASFLVSTIAILGIHYL           613
Qy  434  HQTTL            438
         |||
Db  614  IQITL            618
```

RESULT 38
ABB65548

ID ABB65548 standard; Protein; 652 AA.
AC ABB65548;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 23436.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO20017l042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191639P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL09651.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
PS Disclosure; SEQ ID NO 23436; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL01840-ABL161175), expressed DNA
CC sequences (ABL16176-ABL30511) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 652 AA;

Query Match 7.8%; Score 192; DB 22; Length 652;
Best Local Similarity 21.2%; Pred. No. 1.6e-07;
Matches 103; Conservative 64; Mismatches 154; Indels 164; Gaps 22;

```
Qy   33  VYWNSANKRFQAEG-GYVLYPQIG----------DRLDLLCPRARPPGPHSSPNYBFYKLYLVG   85
Db  219  MHWNTSNSIFRIDNTDHIIDVNKGNLAFEFDQVHIICP-VYEPGTFENET-EKYIIYNVS    276
Qy   86  GAQGRRCEAPPA-PNLLLTCDRPDLDLRFTIKFQEYSPNLWGHEFRSHHDYYIIATSDGT    144
Db  277  KVEYETCRITNADPRVIAICDKPQKLMFFITFRRPFTPQPGGLEFLPGNDYYFISTS--S   334
Qy  145  REGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKPVYSEMPMERDRGAAHSL-----EPG   199
Db  335  KDDLYRRIGGRCSTNNMKVVFKVCCAPEDNN-KTTALSNSKSVTDTGGAINVNIANNDES   393
Qy  200  KENLPGDPTSNAISRGAEG---------------PLPP-----------P            223
Db  394  HVNSHGNNIAIGTNIGINGGQIIGGPQSAGIPINPLSGNNNINGIPTTINSNIDQFNRIP   453
Qy  224  SMPAVAGAAGGLALLLLGVAGAGGAMCWRRRAKPSESRHPGPG-SFGRGGSLGLGGGGG    282
Db  454  IQPNIIGNHVGTNAVGTGIVGGGGIIL------------------TPGHAHGNINMLQPGRGGI   499
Qy  283  MGPREAEPG----ELGIALRG--------------------GGAADPPPCPHYEKVSGDYGH   320
Db  500  NG------AYPGHHHIQTGIRINNVPTQHNYPSHKGNANSNINGNDDH--HHYNK------H     547
Qy  321  PVYIVQD------GPPQSPNIYYTSISVLEWPILHTIQLFFMRSKCSRVTTFLFPVQV     373
Db  374  ITTSTCRMTSFSFTTLNPSMQACRAQMGEFRIRWCFWGDRILGTALFVLVLILLLGRLNM   433
Qy  374  PNEVVRNEELTYNSGAATSDGNIFALWIWLS--------------------IFP--L     583
Db  584  LSIQSCHLSSY--------------WI-------SASFLVSTIAILGIHYL           613
Qy  434  HQTTL            438
Db  614  IQITL            618
```

RESULT 39
AAR82605

ID AAR82605 standard; Protein; 234 AA.
AC AAR82605;
XX
DT 16-MAY-1996 (first entry)
XX
DE Eph transmembrane tyrosine kinase family ligand, Efl-2.
XX
KW Efl-2; EHK1-L; Eph transmembrane tyrosine kinase family ligand;
KW neurological disorder; identification; diagnosis.
XX
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FT Peptide   1..30
FT   /label= signal_peptide

FT   Region   218..235
FT   /note= "carboxy terminal hydrophobic
FT   GPI-recognition tail"
XX
FT   Misc-difference 158
FT   /note= "residue borders main conserved regions"
XX
PD   12-OCT-1995.
XX
PN   WO9527060-A2.
XX
PF   04-APR-1995;   95WO-US04208.
XX
PR   21-OCT-1994;   94US-0327423.
PR   04-APR-1994;   94US-0222075.
PR   12-APR-1994;   94US-0229402.
PR   01-SEP-1994;   94US-0299567.
XX
PA   (REGE-) REGENERON PHARM INC.
XX
PI   Aldrich TH, Davis S, Gale N, Goldfarb M, Maisonpierre PC;
PI   Yancopoulos GD;
XX
DR   WPI; 1995-358635/46.
DR   N-PSDB; AAT03883.
XX
PT   Ligands which bind Eph family receptors - used in the diagnosis of
PT   neurological disorders
XX
PS   Disclosure; Fig 2; 58pp; English.
XX
CC   Efl-2 (also known as EHK-1L) is an Eph transmembrane tyrosine kinase
CC   family ligand. It has homology with B61 (Efl-1) (see AAR82604). Efl-2
CC   ends in a C-terminal hydrophobic sequence that appears to to be a
CC   recognition sequence allowing it to be GPI-linked and thus lacking in
CC   an intracellular domain. Efl-2 is useful for identifying other ligands
CC   for Ehk-1, -2, -3, Eck and Elk receptors. The ligands are useful in
CC   promoting a differential function and/or influencing the phenotye,
CC   such as growth and/or proliferation, of receptor bearing cells. They
CC   may be used in the diagnosis, and treatment of neurological disorders.
XX
SQ   Sequence   234 AA;

Query Match   7.4%; Score 181; DB 16; Length 234;
Best Local Similarity   28.4%; Pred. No. 3.6e-07;
 Matches   64; Conservative 23; Mismatches 82; Indels 56; Gaps 11;

Qy   7 GPGGVRVGALLLLGVLGLIVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDLLCP--R 64
              |||| ||||| |:|: |||||| || || |
Db   24 GPG--------GALG-----NRHAVYWNSSNQHLRRE-GYTVQVNVNDYLDIYCPHYN 67

Qy   65 ARPPGPHSSSNYEFYKLYLYVGGAQGRRCEAPPAPNLLLTCDRPDL---DLRFTIKFQEYS 121
              |||| | |: || : | ||| |||
Db   68 SSGAGPGPGGGAEQYVLYMVSRNGYRTCNASQGFK-RWECNRPHAPHSPIKFSEKFQRYS 126

Qy   122 PNLWGHEFRSHHDYIIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKPV 181
              || | ||| |||| |||||: || | |: |||| |: | |
Db   127 AFSLGYEFHAGHEYYYISTPTHNLH------WKCLR--MKVFVCCASTSHSG-----EKPV 174

Qy   182 SEMP-----------MERDRGAAHSLE------PGKENLP 204
              |:| ||:|| | ||||||
Db   175 PTLPQFTMGPNVKINVLEDFEGENPQVPKLEKSISGTSPKREHLP 219

RESULT 40
AAR71481
AAR71481 standard; Protein; 238 AA.
XX
DT   03-OCT-1995 (first entry)
XX
AC   AAR71481;
XX
DE   Human hek-L protein.
XX

KW   Ligand; cell surface; tyrosine kinase receptor; tumorigenesis;
KW   immunogen.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   Peptide   1..19
FT   /note= "signal peptide"
FT   Protein   20..238
XX
PN   WO9506065-A.
XX
PD   02-MAR-1995.
XX
PF   17-AUG-1994;   94WO-US09282.
XX
PR   20-AUG-1993;   93US-0109745.
PR   30-AUG-1993;   93US-0114426.
PR   03-DEC-1993;   93US-0161132.
PR   09-MAY-1994;   94US-0240124.
XX
PA   (IMMV) IMMUNEX CORP.
XX
PI   Beckmann MP, Cerretti DP;
XX
DR   WPI; 1995-106811/14.
DR   N-PSDB; AAQ85887.
XX
PT   New isolated DNA encoding hek-L protein or its fusion products -
PT   useful as assay reagent or for carrying therapeutic and
PT   diagnostic compounds to leukaemia cells.
XX
PS   Claim 21; Page 36; 45pp; English.
XX
CC   The sequence is that of a novel protein designated hek-L, a protein
CC   that can bind hek (a cell surface receptor tyrosine kinase). Hek-L
CC   is the first known ligand for hek and can be used to study cellular
CC   processes regulated by hek (which may be involved in tumorigenesis).
CC   It is also an immunogen for antibody production, as a reagent for
CC   detecting hek or hek-L in in vitro assays, to determine binding of
CC   hek proteins, to purify hek proteins, and to carry diagnostic or
CC   cytotoxic agents to particular leukaemia cells that express the hek
CC   antigen. Hek-L also binds the elk tyrosine kinase receptor.
CC   See also AAR71482.
XX
SQ   Sequence   238 AA;

Query Match   7.3%; Score 179; DB 16; Length 238;
Best Local Similarity   28.4%; Pred. No. 5.4e-07;
 Matches   65; Conservative 24; Mismatches 80; Indels 60; Gaps 12;

Qy   7 GPGGVRVGALLLLGVLGLIVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDLLCP--R 64
              |||| ||||| |:|: |||||| || || |
Db   24 GPG--------GALG-----NRHAVYWNSSNQHLRRE-GYTVQVNVNDYLDIYCPHYN 67

Qy   65 ARPPGPHSSP----NYEFYKLYLYVGGAQGRRCEAPPAPNLLLTCDRPDL---DLRFTIKF 117
              |||| | |: || : | ||| ||| |||
Db   68 SSGVGPGAGPGPGGGAEQYVLYMVSRNGYRTCNASQGFK-RWECNRPHAPHSPIKFSEKF 126

Qy   118 QEYSPNLWGHEFRSHHDYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVP 177
              |:|| |||| |||| |||||: || | |: |||| |: | |
Db   127 QRYSEAFSLGYEFHAGHEYYYISTPTHNLH------WKCLR--MKVFVCCASTSHSG--- 174

Qy   178 RKPVSEMP-----------MERDRGAAHSLE------PGKENLP 204
              |:|| ||:|| | ||||||
Db   175 EKPVPTLPQFTMGPNVKINVLEDFEGENPQVPKLEKSISGTSPKREHLP 223

RESULT 41
AAW02587
AAW02587 standard; Protein; 184 AA.
XX
ID   AAW02587
XX
AC   AAW02587;

XX
DT   28-NOV-1996 (first entry)
XX
DE   Lerk-6 protein.
XX
KW   Lerk-6; hek; elk; cell surface receptor; culture; reagent;
KW   neuron; disorder; injury; delivery agent; diagnostic; therapeutic.
KW   Lerk-7; probe; cytokine.
XX
OS   Mus musculus.
XX
PN   WO9617925-A1.
XX
PD   13-JUN-1996.
XX
PF   05-DEC-1995;   95WO-US15781.
XX
PR   01-MAR-1995;   95US-0396946.
PR   06-DEC-1994;   94US-0351025.
XX
PA   (IMMV) IMMUNEX CORP.
XX
PI   Cerretti DP;
XX
DR   WPI; 1996-281171/29.
DR   N-PSDB; AAT32700.
XX
PT   New isolated human Lerk-7 cytokine - which binds to cell surface
PT   receptors elk, hek and eck, useful for delivering agents to cells or
PT   for treating neural disorders
XX
PS   Example 1; Page 35; 49pp; English.
XX
CC   The present sequence is that of a murine cytokine, Lerk-6, encoded by
CC   AAT32700. A fragment of murine Lerk-6 DNA was isolated by PCR for use as
CC   a probe for a human Lerk-6 gene (see AAT32699). The probe however led to the
CC   identification of a human Lerk-6 homologue. The Lerk-7 gene
CC   and protein can be used for studying the role of Lerk-7 in conjunction
CC   with elk, hek and eck receptors. They can also be used for delivering
CC   diagnostic or therapeutic agents to cells, e.g. cancer cells. The
CC   Lerk-7 proteins can also exhibit neuroprotective or neurotrophic
CC   properties and can be used to treat neural tissue disorders.
XX
SQ   Sequence 184 AA;

Query Match   7.2%; Score 176; DB 17; Length 184;
Best Local Similarity   29.3%; Pred. No. 7e-07;
Matches   58; Conservative 19; Mismatches 69; Indels 52; Gaps 7;

Qy   33 VYWNSANKFQAE-----GGYVLYPQIGDRLDLLCPRARPPGPHSSPNYEFYKLYLVGGA 87
         |||| |:|| |: : | || | | | ::|| |:|| || ||
Db   10 VYWNRSNPRFQVSAVGDGGGYTVEVSINDYLDIYCPHYGAPLP-PAERMERYILYMVNGE 68

Qy   88 QGRRCEAPPAPNLLLTCDRPDL---DLRFTIKFQEYSPNIWGHEFRSHHDYIIIATSDGT 144
         |:|| | || |||:|||||| |: |::| : || :: |: ||: || | |
Db   69 GHASCDHRQRGFKRWECNRPAAPGGPLKFSEKFQLFTTPFSLGFEFRPGHEYYYISATP-- 126

Qy   145 REGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKPVSEMPERDRGAAHSLEPGKENLP 204
         ::   : :: |:
Db   127 ----PNLVDRPCLR--LKYYVR-------------------------PTNETLY 149

Qy   205 GDP----TSNATSRGAEG 218
         :: |:|| ::
Db   150 EAPEPIFTSNSSCSGLGG 167

RESULT 42
AAR91283
ID   AAR91283 standard; Protein; 184 AA.
XX
AC   AAR91283;
XX
DT   10-OCT-1996 (first entry)

XX
DE   LERK-6 protein.
XX
KW   LERK-6; hek; elk; cell surface receptors; culture; reagent;
KW   neurons; disorder; injury; delivery agent; diagnostic; therapeutic.
XX
OS   Mus musculus.
XX
PN   WO9610911-A1.
XX
PD   18-APR-1996.
XX
PF   04-OCT-1995;   95WO-US12779.
XX
PR   03-OCT-1995;   95US-0538709.
PR   05-OCT-1994;   94US-0318393.
XX
PA   (IMMV) IMMUNEX CORP.
XX
PI   Cerretti DP;
XX
DR   WPI; 1996-209575/21.
DR   N-PSDB; AAT14009.
XX
PT   Isolated DNA encoding cytokine designated LERK-6 which binds to hek
PT   and elk cell surface receptors - useful for drug delivery and
PT   screening procedures.
XX
PS   Claim 6; Page 34; 44pp; English.
XX
CC   The LERK-6 polypeptide encoded can be used to isolate cells
CC   expressing hek/elk cell surface receptors, or to measure the
CC   biological activity of such receptors. The protein may also be used
CC   as a delivery agent, taking diagnostic and therapeutic agents to
CC   cells expressing such receptors. LERK-6 can also be used as a tissue
CC   culture reagent to enhance the viability or prolong the lifespan of
CC   the neurons. Neural tissue disorders and injuries may be treated by
CC   contact with the polypeptide.
XX
SQ   Sequence 184 AA;

Query Match   7.2%; Score 176; DB 17; Length 184;
Best Local Similarity   29.3%; Pred. No. 7e-07;
Matches   58; Conservative 19; Mismatches 69; Indels 52; Gaps 7;

Qy   33 VYWNSANKFQAE-----GGYVLYPQIGDRLDLLCPRARPPGPHSSPNYEFYKLYLVGGA 87
         |||| |:|| |: : | || | | | ::|| |:|| || ||
Db   10 VYWNRSNPRFQVSAVGDGGGYTVEVSINDYLDIYCPHYGAPLP-PAERMERYILYMVNGE 68

Qy   88 QGRRCEAPPAPNLLLTCDRPDL---DLRFTIKFQEYSPNIWGHEFRSHHDYIIIATSDGT 144
         |:|| | || |||:|||||| |: |::| : || :: |: ||: || | |
Db   69 GHASCDHRQRGFKRWECNRPAAPGGPLKFSEKFQLFTTPFSLGFEFRPGHEYYYISATP-- 126

Qy   145 REGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKPVSEMPERDRGAAHSLEPGKENLP 204
         ::   : :: |:
Db   127 ----PNLVDRPCLR--LKYYVR-------------------------PTNETLY 149

Qy   205 GDP----TSNATSRGAEG 218
         :: |:|| ::
Db   150 EAPEPIFTSNSSCSGLGG 167

RESULT 43
AAY06820
ID   AAY06820 standard; Protein; 184 AA.
XX
AC   AAY06820;
XX
DT   24-JUN-1999 (first entry)
XX
DE   Murine LERK-6 polypeptide.
XX
KW   LERK-6 polypeptide; hek receptor; elk receptor; human; murine; injury;

KW cell proliferation; neural growth; neural tissue; neurological disease;
KW neurodegenerative; excitotoxicity.
XX
OS Mus sp.
XX
PN WO9910495-A1.
XX
PD 04-MAR-1999.
XX
PF 27-AUG-1998; 98WO-US17772.
XX
PR 29-AUG-1997; 97US-0920440.
XX
PA (IMMV) IMMUNEX CORP.
XX
PI Cerretti DP;
XX
DR WPI; 1999-243567/20.
DR N-PSDB; AAX32761.
XX
PT New cytokine designated LERK-6
XX
PS Claim 6; Page 37-38; 46pp; English.
XX
CC The invention relates murine and human LERK-6 polypeptides that bind to
CC hek/elk receptors. Host cells transfected or transformed with vectors
CC comprising the LERK-6 nucleic acid sequences are used for the recombinant
CC production of the proteins. LERK-6 polypeptides may be useful in the
CC enhancement, stimulation, proliferation or growth of cells expressing the
CC hek or elk receptor. The ligand and receptor complex may be involved in
CC neural growth, development and/or maintenance. LERK-6 can be used for
CC treating disorders of neural tissue such as injury or neurological
CC diseases, either chronic or acute. LERK-6 may be employed in treating
CC neurodegenerative conditions where there is neural death, excitotoxicity.
CC In addition, they may be administered to a mammal to exert a trophic
CC effect on neural tissue. They can also be used as reagents for those
CC conducting quality assurance studies e.g. to monitor shelf life and
CC stability of elk protein under different conditions. The polypeptides can
CC also be used as carriers for delivering agents attached to cells bearing
CC the elk or hek cell surface receptor. The present sequence represents a
CC murine LERK-6 polypeptide.
XX
SQ Sequence   184 AA;

Query Match   7.2%; Score 176; DB 20; Length 184;
Best Local Similarity   29.3%; Pred. No. 7e-07;
Matches 58; Conservative 19; Mismatches 69; Indels 52; Gaps   7;

Qy   33 VYWNSANKRFQAE------GGYVLYPQIGDRLDLLCPRARPPGPHSSPNYEFYKLYLVGGA 87
                         ::|||||||||||||||||||||||||||||||||||||||||:
Db   10 VYWNRSNPRFQVSAVGDGGGYTVEVSINDYLDIYCPHYGAPLP-PAERMERYILYMVNGE 68

Qy   88 QGRCEAPPAPNLLLTCDRPDL---DLRFTIKFQEYSPNLWGHEFRSHHDYYIIATSDGT 144
                |:   :|: |||||||||||| : :|:||||||:||||||||:
Db   69 GHASCDHRQRGFKRWECNRPAAPGGPLKFSEKFQLFTPFSLGFEFRPGHEYYYISATP-- 126

Qy   145 REGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKPVSEMPMERDRGAAHSLEPGKENLP 204
                ||:|:||| :|||||||||||||||||||:|:||||:
Db   127 ----PNLVDRPCLR--LKVYVR------------------------------PTNETLY 149

Qy   205 GDP----TSNATSRGAEG 218
                :|:
Db   150 EAPEPIFTSNSSCSGLGG 167

DE Mouse EPH receptor ligand Elf-1.
XX
KW Elf-1; EPH receptor ligand; dementia; tachycardia; therapy;
KW diagnosis; transgenic animal.
XX
OS Mus musculus.
XX
FH Key   Location/Qualifiers
FT Peptide   1..20
FT   /label= Sig_peptide
FT Protein   21..209
FT   /label= Mat_protein
FT Region   69..159
FT   /label= Cys4_motif
FT Region   35..166
FT   /label= Core_sequence_motif
FT Modified-site 38
FT   /note= "potential N-glycosylation site"
FT Modified-site 170
FT   /note= "potential N-glycosylation site"
FT Modified-site 184
FT   /note= "potential N-glycosylation site"
XX
PN WO9609384-A1.
XX
PD 28-MAR-1996.
XX
PF 19-SEP-1995; 95WO-US11869.
XX
PR 27-FEB-1995; 95US-0393462.
PR 19-SEP-1994; 94US-0308814.
XX
PA (HARD) HARVARD COLLEGE.
XX
PI Cheng H, Flanagan JG;
XX
DR WPI; 1996-184446/19.
DR N-PSDB; AAT15008.
XX
PT Murine and chicken EPH receptor ligand, Elf-1 - useful in diagnosis
PT and treatment of disorders associated with the Elf-1 gene, e.g.
PT dementia, tachycardia , etc.
XX
PS Claim 1; Page 86; 107pp; English.
XX
CC A novel mouse EPH receptor ligand, Elf-1 (AAR94766), is involved in
CC the formation and maintenance of ordered spatial arrangements of
CC differentiated tissue. It is the product of a cDNA clone (AAT15008)
CC obtd. from an embryo mid- and hind-brain cDNA expression library.
CC This cDNA is used for the prodn. of recombinant Elf-1, which can
CC be used to modulate proliferation, survival and/or differentiation
CC of cells and tissues, and to stimulate or antagonise intracellular
CC signal transduction pathways mediated by the EPH-type receptor.
XX
SQ Sequence   209 AA;

Query Match   7.2%; Score 176; DB 17; Length 209;
Best Local Similarity   29.3%; Pred. No. 8.2e-07;
Matches 58; Conservative 19; Mismatches 69; Indels 52; Gaps   7;

Qy   33 VYWNSANKRFQAE------GGYVLYPQIGDRLDLLCPRARPPGPHSSPNYEFYKLYLVGGA 87
                         ::|||||||||||||||||||||||||||||||||||||||||:
Db   35 VYWNRSNPRFQVSAVGDGGGYTVEVSINDYLDIYCPHYGAPLP-PAERMERYILYMVNGE 93

Qy   88 QGRCEAPPAPNLLLTCDRPDL---DLRFTIKFQEYSPNLWGHEFRSHHDYYIIATSDGT 144
                |:   :|: |||||||||||| : :|:||||||:||||||||:
Db   94 GHASCDHRQRGFKRWECNRPAAPGGPLKFSEKFQLFTPFSLGFEFRPGHEYYYISATP-- 151

Qy   145 REGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKPVSEMPMERDRGAAHSLEPGKENLP 204
                ||:|:||| :|||||||||||||||||||:|:||||:
Db   152 ----PNLVDRPCLR--LKVYVR------------------------------PTNETLY 174

Qy   205 GDP----TSNATSRGAEG 218

ID AAR94766 standard; Protein; 209 AA.
XX
AC AAR94766;
XX
DT 02-JUL-1996 (first entry)
XX

RESULT 44
AAR94766

Db   175 EAPEPIFTSNSSCSGLGG 192

RESULT 45

AAW71006
ID   AAW71006 standard; Protein; 209 AA.
XX
AC   AAW71006;
XX
DT   20-OCT-1998 (first entry)
XX
DE   Amino acid sequence of a mammalian Elf-1 protein.
XX
KW   Mouse; EPH receptor ligand; Elf-1; mek-4; sek-AP; tyrosine kinase ligand;
KW   B61; LERK-2; proliferation; differentiation; intracellular signalling;
KW   increased; survival; neuronal cell; neuron survival; treatment;
KW   Alzheimer's; Parkinson's; lymphatic tumour; artificial liver; cartilage;
KW   bone formation.
XX
OS   Mus sp.
XX
FH   Key   Location/Qualifiers
FH
FT   Peptide   1..20
FT   /note="signal peptide"
FT   Protein   21..209
FT   /note="mature protein"
FT   Modified-site   38
FT   /note="potential N-linked glycosylation site"
FT   Modified-site   35..166
FT   /note="contains core sequence motif"
FT   Misc-difference 69..159
FT   /note="contains a Cys4 motif"
FT   Modified-site   170
FT   /note="potential N-linked glycosylation site"
FT   Modified-site   184
FT   /note="potential N-linked glycosylation site"
XX
XX
PN   US5795734-A.
XX
PD   18-AUG-1998.
XX
PF   31-MAY-1995; 95US-0455001.
XX
PR   31-MAY-1995; 95US-0455001.
PR   19-SEP-1994; 94US-0308814.
PR   27-FEB-1995; 95US-0393462.
XX
PT   Nucleic acid encoding Elf-1 protein that binds to EPH-type receptor
PT   - for production of Elf-1 protein, useful for regulating
PT   proliferation, differentiation, and survival of cells
XX
PA   (HARD) HARVARD COLLEGE.
XX
PI   Cheng H, Flanagan JG;
XX
DR   WPI; 1998-466665/40.
DR   N-PSDB; AAV42926.
XX
PS   Claim 1; Fig 2A; 5pp; English.
XX
CC   The present sequence represents a mammalian EPH receptor ligand
CC   designated Elf-1. This ligand can bind to both mek-4 and sek-AP. Elf-1 is
CC   a tyrosine kinase ligand, which is linked to the membrane through a
CC   phosphatidylinositol linkage. It shares some homology to 2 other EPH
CC   receptor ligands, B61 and LERK-2. The Elf-1 protein modulates
CC   proliferation, differentiation and survival of EPH receptor-expressing
CC   cells by stimulating or antagonising intracellular signalling mediated
CC   by the EPH receptor. Typical of many potential applications are
CC   increasing survival of neuronal cells in culture (e.g. where intended for
CC   transplantation), also therapeutically in increase neuron survival
CC   (e.g. treatment of Alzheimer's or Parkinson's diseases), to prevent
CC   nervous system and lymphatic tumours, to induce differentiation of

CC   hepatocytes to form an artificial liver, to induce cartilage and bone
CC   formation.
XX
SQ   Sequence 209 AA;

Qy   33 VYWNSANRFQAB-----GGGYIYPQIGDRLDLLCPRARPGPHSSPNYEFYKLIYLVGGA 87
        |||::|||||.| ||:|..:|||...|.|.||.|:.::||.|:.|.|||
Db   35 VYWNRSNPRFQVSAVGDGGGITVEVSINDYLDIYCPHYGAPLP--PAERMERYILYMVNGE 93

Qy   88 QGRRCEAPPAPNLLLTCDRPDL---DLRFT1KFQEYSPNLWGHEFRSHHDYYIIATSDGT 144
        .|||.|||.|.|||:|:.|:|| |:|||:|||:|||.||.|:|||||::..||
Db   94 GHASCDHRQRGFKRWECNRPAAPGGPLKFSEKFQLFTPFSLGFEFRPGHEYYISATP-- 151

Qy 145 REGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKPVSEMPMERDRGAAHSLEPGKENLP 204
        |.:|.:|..||||.:|..||:||::|:|..||:|:|.:||.::|.||.|:|||.||:|||
Db 152 ----PNLVDRPCLR--LKVYVR------------------------------PTNETLY 174

Qy 205 GDP-----TSNAISRGAEG 218
            .| ||:::|||.|
Db 152 EAPEPIFTSNSSCSGLGG 192

RESULT 46

AAY06822
ID   AAY06822 standard; Protein; 213 AA.
XX
AC   AAY06822;
XX
DT   24-JUN-1999 (first entry)
XX
DE   Human LERK-6 polypeptide.
XX
KW   LERK-6 polypeptide; hek receptor; elk receptor; human; murine; injury;
KW   cell proliferation; neural growth; neural tissue; neurological disease;
KW   neurodegenerative; excitotoxicity.
XX
OS   Homo sapiens.
XX
PN   WO9910495-A1.
XX
PD   04-MAR-1999.
XX
PI   Cerretti DP;
XX
PA   (IMMV) IMMUNEX CORP.
XX
PR   29-AUG-1997; 97US-0920440.
XX
PF   27-AUG-1998; 98WO-US177772.
XX
DR   WPI; 1999-243567/20.
DR   N-PSDB; AAX32767.
XX
PT   New cytokine designated LERK-6
XX
PS   Claim 6; Page 42; 46pp; English.
XX
CC   The invention relates murine and human LERK-6 polypeptides that bind to
CC   hek/elk receptor. Host cells transfected or transformed with vectors
CC   comprising the LERK-6 nucleic acid sequences are used for the recombinant
CC   production of the proteins. LERK-6 nucleic acid sequences may be useful in the
CC   enhancement, stimulation, proliferation or growth of cells expressing the
CC   hek or elk receptor. The ligand and receptor complex may be involved in
CC   neural growth, development and/or maintenance. LERK-6 can be used for
CC   treating disorders of neural tissue such as injury or neurological
CC   disease, either chronic or acute. LERK-6 may be employed in treating
CC   neurodegenerative conditions where there is neural death, excitotoxicity.
CC   In addition, they may be administered to a mammal to exert a trophic
CC   effect on neural tissue. They can also be used as reagents for those

```
CC   conducting quality assurance studies e.g. to monitor shelf life and
CC   stability of elk protein under different conditions. The polypeptides can
CC   also be used as carriers for delivering agents attached to cells bearing
CC   the elk or hek cell surface receptor. The present sequence represents a
CC   human LERK-6 polypeptide.
XX
SQ   Sequence   213 AA;

Query Match             7.2%;  Score 175.5;  DB 20;  Length 213;
Best Local Similarity  36.8%;  Pred. No. 9.3e-07;
Matches   43;  Conservative  14;  Mismatches  51;  Indels  9;  Gaps  3;

Qy          33 VYWNSANKRFQA-----EGGYVLYPQIGDRLDLLCPRARPPGPHSSPNYEFYKLYLVGGA 87
                |||:|:||:||       ||| || :||| | :|| | ||| |:|| ||:||||||||
Db          39 VYWNRSNPRFHAGAGDDGGGYTVEVSINDYLDIYCPHYGAPLP-PAERMEHYLYMVNGE 97

Qy          88 QGRRCEAPPAPNLLLTCDRPDL---DLRFIKFQEYSPNIWGHEFRSHHDYIIATS 141
                |:| |||::|  :|:|||||    :|||::| |||  |||| |:||| :||:
Db          98 GHASCDHRQGPKRWECNRPAAPGGPLKFSEKFQLFTPFSLGFPFRPGHEYYISAT 154

RESULT 47
AAB54058
ID   AAB54058 standard; Protein; 218 AA.
XX
AC   AAB54058;
XX
DT   09-MAR-2001 (first entry)
XX
DE   Human pancreatic cancer antigen protein sequence SEQ ID NO:510.
XX
KW   Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
KW   detection; diagnosis; identification; cytostatic; neuroprotective;
KW   nootropic; immunomodulatory; relaxant; contraceptive; gynaecological;
KW   antiinflammatory; cardiant; gene therapy; chromosome mapping;
KW   linkage analysis; tissue identification; tissue typing; forensic;
KW   neural; immune system; muscular; reproductive; gastrointestinal;
KW   pulmonary; cardiovascular; renal; proliferative.
XX
OS   Homo sapiens.
XX
PN   WO20005320-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05989.
XX
PR   12-MAR-1999;    99US-0124270.
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2000-579444/54.
DR   N-PSDB; AAC98823.
XX
PT   New nucleic acid that is a pancreatic cancer antigen for preventing,
PT   treating, or ameliorating a medical condition, particular pancreatic
PT   cancer, or for use in assays for diagnosing a pathological condition -
PS   Claim 11; Page 946-947; 1379pp; English.
XX
CC   AAC98773 to AAC9231 encode the human pancreatic cancer associated
CC   proteins, called pancreatic cancer antigens, given in AAB54008 to
CC   AAB54466. The human pancreatic cancer antigens have cytostatic,
CC   neuroprotective, nootropic, immunomodulatory, relaxant, contraceptive,
CC   gynaecological, cardiant and antiinflammatory activities, and can be used
CC   in gene therapy. The polynucleotide and proteins can be used for
CC   preventing, treating, or ameliorating a medical condition or in assays
CC   for diagnosing a pathological condition or a susceptibility to one in a
CC   subject. Binding partners to the proteins and the activity of the
CC   proteins can be identified. The pancreatic cancer antigens can be used to
CC   detect, treat or prevent pancreatic disorders, especially cancer.
```

```
CC   Agonists and antagonists to the antigens can be screened for. The
CC   pancreatic cancer antigen polynucleotides can be used to design nucleic
CC   acid hybridisation probes that can be used in chromosome mapping, linkage
CC   analysis, tissue identification and/or typing and a variety of forensic
CC   and diagnostic methods. The proteins can be used to generate antibodies
CC   which are used to purify, detect and target the polypeptides, including
CC   both in vivo and in vitro diagnostic and therapeutic methods. The
CC   proteins can be used to treat or prevent neural, immune system, muscular,
CC   reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
CC   proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence   218 AA;

Query Match             7.0%;  Score 172.5;  DB 21;  Length 218;
Best Local Similarity  28.3%;  Pred. No. 1.7e-06;
Matches   51;  Conservative  30;  Mismatches  80;  Indels  19;  Gaps  6;

Qy           8 PGGVRVGALILLGVLGLVSGLSL--EPVYWNSANKRFQAEGGYVLYPQIGDRLDLLCPR 64
                || | : ::|| ::|:||| :|    |:|||:|:|||:| :||| || :||| | :||
Db           8 PGQARAMEFLWAPLLGLCCSLAAADRHTVFWNSSNPKFRNE-DYTIHVQLNDYVDICPH 66

Qy          65 ARPPGPHSSPN--YEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDL---DLRFTIKPQE 119
                   | : |:     |:||:|||: :| | |:| |||::|   |||    :|||::| |
Db          67 YE---DHSVADAAMEQYILYLVEHEEYQLCQPQSKDQVRWQCNRPSAKGHGPEKLSEKFQR 123

Qy         120 YSPNLWGHEFRSHHDYIYIIATSDGTRGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRK 179
                ::| |||| |:||| ::|| ||| | ||  |        |:|  | |  :|   ||: |:
Db         124 FTPFTLGKFFKEGHSYYISKPIHQHEDR------CLRLKVTVSGKITHSPQAHDNPQE 176

RESULT 48
AAR94767
ID   AAR94767 standard; Protein; 200 AA.
XX
AC   AAR94767;
XX
DT   02-JUL-1996 (first entry)
XX
DE   Chicken EPH receptor ligand Elf-1.
XX
KW   Elf-1; EPH receptor ligand; dementia; tachycardia; therapy;
KW   diagnosis; transgenic animal.
XX
OS   Gallus sp.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..21
FT                   /label= Sig_peptide
FT   Protein         22..200
FT                   /label= Mat_protein
FT   Region          61..150
FT                   /label= Cys4_motif
FT   Region          33..157
FT                   /label= Core_sequence_motif
XX
PN   WO9609384-A1.
XX
PD   28-MAR-1996.
XX
PF   19-SEP-1995; 95WO-US11869.
XX
PR   27-FEB-1995;    95US-0393462.
PR   19-SEP-1994;    94US-0308814.
PA   (HARD ) HARVARD COLLEGE.
XX
PI   Cheng H,  Flanagan JG;
XX
DR   WPI; 1996-184446/19.
DR   N-PSDB; AAT15009.
XX
PT   Murine and chicken EPH receptor ligand, Elf-1 - useful in diagnosis
```

```
PT   and treatment of disorders associated with the Elf-1 gene, e.g.
PT   dementia, tachycardia , etc.
XX
PS   Claim 1; Page 88-89; 107pp; English.
XX
CC   A novel chicken EPH receptor ligand, Elf-1 (AAR94767), is involved in
CC   the formation and maintenance of ordered spatial arrangements of
CC   differentiated tissue. It is the product of a cDNA clone (AAT15009)
CC   obtd. from an embryo cDNA expression library. This cDNA is used for
CC   the prodn. of recombinant Elf-1, which can be used to modulate
CC   proliferation, survival and/or differentiation of cells and tissues,
CC   and to stimulate or antagonise intracellular signal transduction
CC   pathways mediated by the EPH-type receptor.
XX
SQ   Sequence   200 AA;

Query Match    7.0%;    Score 172;  DB 17;  Length 200;
Best Local Similarity    35.3%;  Pred. No. 1.7e-06;
Matches   49;  Conservative   16;  Mismatches   58;  Indels   16;  Gaps   5;

Qy    15  ALILLGVLGLVSGLSLEP-----------------VYWNSANKRFQAEGGYVLYPQIGDRLDLLCPRA   65
          ::|:|||||||||||::|                   :::|:::|:|||||||:|:|||  :|:::::|
Db     7  AALLAAIVG-VCVWSDDPGKVISDRYAVYWNRSNPRFH-RGDYTVEVSINDYLDIYCPHY   64

Qy    66  RPPGPHSSPNYEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDL----DLRFTIKFQEYSP   122
          |||||  |:::|||||| ||:|  |  :    ||::| ||  |:    :|| ||| |:||
Db    65  EEPLP--AERMERYVLYMVNYEGHASCDHRQKGFKRWECNRPDSPSGPLKFSEKFQLFTP   122

Qy   123  NLWGHEFRSHHDYYIIATS   141
          ||| |||| ::|||:
Db   123  FSLGFEFRPGHEYYISAS   141

RESULT 49
AAW71007
ID   AAW71007 standard; Protein; 200 AA.
XX
AC   AAW71007;
XX
DT   20-OCT-1998  (first entry)
XX
DE   Amino acid sequence of an avian Elf-1 protein.
XX
KW   Chicken; EPH receptor ligand; Elf-1; mek-4; sek-AP;
KW   tyrosine kinase ligand; B61; LERK-2; proliferation; differentiation;
KW   intracellular signalling; increased; survival; neuronal cell;
KW   neuron survival; treatment; Alzheimer's; Parkinson's; lymphatic tumour;
KW   artificial liver; cartilage; bone formation.
XX
OS   Gallus sp.
XX
FH   Key              Location/Qualifiers
FH
FT   Misc-difference  61..150
FT                    /note= "contains a Cys4 motif"
FT   Misc-difference  35..157
FT                    /note= "contains a core sequence motif"
XX
PN   US5795734-A.
XX
PD   18-AUG-1998.
XX
PF   31-MAY-1995;   95US-0455001.
XX
PR   31-MAY-1995;   95US-0455001.
PR   19-SEP-1994;   94US-0308814.
PR   27-FEB-1995;   95US-0393462.
XX
PA   (HARD ) HARVARD COLLEGE.
XX
PI   Cheng H,  Flanagan JG.
XX
DR   WPI; 1998-466665/40.
DR   N-PSDB; AAV42927.
```

```
XX
PT   Nucleic acid encoding Elf-1 protein that binds to EPH-type receptor
PT   - for production of Elf-1 protein, useful for regulating
PT   proliferation, differentiation, and survival of cells
XX
PS   Claim 1; Columns 75-78; 53pp; English.
XX
CC   The present sequence represents an avian EPH receptor ligand designated
CC   Elf-1. This ligand can bind to both mek-4 and sek-AP. Elf-1 is a
CC   tyrosine kinase ligand, which is linked to the membrane through a
CC   phosphatidylinositol linkage. It shares some homology to 2 other EPH
CC   receptor ligands, B61 and LERK-2. The Elf-1 protein modulates
CC   proliferation, differentiation and survival of EPH receptor-expressing
CC   cells by stimulating or antagonising intracellular signalling mediated
CC   by the EPH receptor. Typical of many potential applications are
CC   increasing survival of neuronal cells in culture (e.g. where intended for
CC   transplantation), also therapeutically in increase neuron survival
CC   (e.g. treatment of Alzheimer's or Parkinson's diseases), to prevent
CC   nervous system and lymphatic tumours, to induce differentiation of
CC   hepatocytes to form an artificial liver, to induce cartilage and bone
CC   formation.
XX
SQ   Sequence   200 AA;

Query Match    7.0%;    Score 172;  DB 19;  Length 200;
Best Local Similarity    35.3%;  Pred. No. 1.7e-06;
Matches   49;  Conservative   16;  Mismatches   58;  Indels   16;  Gaps   5;

Qy    15  ALILLGVLGLVSGLSLEP-----------------VYWNSANKRFQAEGGYVLYPQIGDRLDLLCPRA   65
          ::|:|||||||||||::|                   :::|:::|:|||||||:|:|||  :|:::::|
Db     7  AALLAAIVG-VCVWSDDPGKVISDRYAVYWNRSNPRFH-RGDYTVEVSINDYLDIYCPHY   64

Qy    66  RPPGPHSSPNYEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDL----DLRFTIKFQEYSP   122
          |||||  |:::|||||| ||:|  |  :    ||::| ||  |:    :|| ||| |:||
Db    65  EEPLP--AERMERYVLYMVNYEGHASCDHRQKGFKRWECNRPDSPSGPLKFSEKFQLFTP   122

Qy   123  NLWGHEFRSHHDYYIIATS   141
          ||| |||| ::|||:
Db   123  FSLGFEFRPGHEYYISAS   141

RESULT 50
AAR71482
ID   AAR71482 standard; Protein; 201 AA.
XX
AC   AAR71482;
XX
DT   03-OCT-1995  (first entry)
XX
DE   Human hek-L protein.
XX
KW   Ligand; cell surface; tyrosine kinase receptor; tumorigenesis;
KW   immunogen.
XX
OS   Homo sapiens.
XX
FH   Key              Location/Qualifiers
FT   Peptide          1..22
FT                    /note= "signal peptide"
FT   Protein          23..201
XX
PN   WO9506065-A.
XX
PD   02-MAR-1995.
XX
PF   17-AUG-1994;  94WO-US09282.
XX
PR   20-AUG-1993;  93US-0109745.
PR   30-AUG-1993;  93US-0114426.
PR   03-DEC-1993;  93US-0161132.
PR   09-MAY-1994;  94US-0240124.
XX
PA   (IMMV ) IMMUNEX CORP.
```

XX
PI   Beckmann MP, Cerretti DP;
XX
DR   WPI; 1995-106811/14.
DR   N-PSDB; AAQ85888.
XX
PT   New isolated DNA encoding hek-L protein or its fusion products -
PT   useful as assay reagent or for carrying therapeutic and
PT   diagnostic compounds to leukaemia cells.
XX
PS   Claim 21; Page 38; 45pp; English.
XX
CC   The sequence is that of a novel protein designated hek-L, a protein
CC   that can bind hek (a cell surface receptor tyrosine kinase). Hek-L
CC   is the first known ligand for hek and can be used to study cellular
CC   processes regulated by hek (which may be involved in tumorigenesis).
CC   It is also an immunogen for antibody production, as a reagent for
CC   detecting hek or hek-L in in vitro assays, to determine binding of
CC   hek proteins, to purify hek proteins, and to carry diagnostic or
CC   cytotoxic agents to particular leukaemia cells that express the hek
CC   antigen. Hek-L also binds the elk tyrosine kinase receptors.
CC   See also AAR71481.
XX
SQ   Sequence   201 AA;

Query Match   6.9%;   Score 169.5; DB 16;   Length 201;
Best Local Similarity   29.9%;   Pred. No. 2.8e-06;
Matches   66; Conservative   18; Mismatches   82; Indels   55; Gaps   10;

Qy   24 LVSGLSI-EPYYWNSANKRFQAEGGTVLYPQIGDRLDLLCPRARPPGPHSSPNYEFYKIY 82
          : |||||| :   : |||:||:|: ||| ||:| |||:| :| ||
Db   20 LRGGSSLRHVVYWNSSNPRL-LRGDAVVELGLNDYLDIVCPHYEGPGPEGP--ETFALY 76

Qy   83 LVGGAQGRRCEAP-PAPNLLITCDRPDLDLRFTIKFQEYSPNLWGHEFRSHHDYYIIATS 141
          | | :| | | | |||||: :|:|:|||:| :|||||||||:||: |||||||
Db   77 MVDWPGYESCQAEGPRAYKRWVCSLPFGHVQFSEKIQRPTPFSLGFEFLPGETYYISVP 136

Qy   142 DGTREGLESLQGGVCLTRGMKVLRVGQSPRGGAVPRKPVVSEMPMERDGAAHSLEPGKE 201
          | | |||| |||||:||: | | :|| || || | :||| || ||:|||:|
Db   137 --TPE------SSGQCL----------------------------RLQVSVCCKERKSESAHPV----- 164

Qy   202 NLPGDPTSNATS--RGAEGPLPPPSMPAVAGAAGGLALLLL 240
          : || :: || : : | ||||||
Db   165 ---GSPGESGTSGWRGGDTPSP--------------LCLLLL 189

Search completed: February 11, 2003, 12:02:56
Job time : 43.7736 secs

OM protein - protein search, using sw model

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: February 11, 2003, 12:01:50 ; Search time 14.3082 seconds
(without alignments)
935.649 Million cell updates/sec

Title: US-10-021-121-2
Perfect score: 2450
Sequence: 1 MGPPHSGPGGVRVGALLLLG.............TTLLRQPASVEAEAGQHGPL 455

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1837	75.0	340	4	US-09-214-631-3	Sequence 3, Appli
2	628.5	25.7	336	4	US-09-173-133-2	Sequence 2, Appli
3	628.5	25.7	336	2	US-09-165-533-2	Sequence 2, Appli
4	628.5	25.7	336	5	US-08-436-054-2	Sequence 2, Appli
5	623	25.4	346	1	US-08-213-403-2	Sequence 2, Appli
6	623	25.4	346	1	US-08-458-077-2	Sequence 2, Appli
7	623	25.4	346	1	US-08-460-741-2	Sequence 2, Appli
8	623	25.4	346	1	US-08-747-240-2	Sequence 2, Appli
9	623	25.4	346	1	US-08-299-567-6	Sequence 2, Appli
10	620.5	25.3	333	1	US-08-436-044-4	Sequence 6, Appli
11	620.5	25.3	333	2	US-08-436-054-4	Sequence 4, Appli
12	620.5	25.3	333	4	US-08-271-948-2	Sequence 2, Appli
13	620.5	25.3	333	5	PCT-US95-08534-2	Sequence 2, Appli
14	620.5	25.3	333	5	PCT-US95-08812-4	Sequence 4, Appli
15	613.5	25.0	333	4	PCT-US95-08812-4	Sequence 4, Appli
16	599.5	24.5	345	4	US-09-214-631-5	Sequence 5, Appli
17	489	20.0	89	4	US-09-214-631-13	Sequence 13, Appli
18	192	7.8	82	4	US-09-214-631-11	Sequence 11, Appl
19	191.5	7.8	82	4	US-09-214-631-12	Sequence 12, Appli
20	179	7.3	234	4	US-08-299-567-5	Sequence 5, Appli
21	179	7.3	238	1	US-08-240-124-2	Sequence 2, Appli
22	179	7.3	238	2	US-08-453-943-2	Sequence 2, Appli
23	179	7.3	238	2	US-09-057-121-2	Sequence 2, Appli
24	179	7.3	238	4	US-09-358-734-2	Sequence 2, Appli
25	176.5	7.2	135	1	US-08-299-567-7	Sequence 7, Appli
26	176	7.2	184	1	US-08-609-324A-2	Sequence 2, Appli
27	176	7.2	184	2	US-09-609-440B-2	Sequence 2, Appli
28	176	7.2	184	4	US-09-173-492-2	Sequence 2, Appli
29	176	7.2	184	4	US-09-173-133-2	Sequence 2, Appli
30	176	7.2	184	4	US-09-165-533-2	Sequence 2, Appli
31	176	7.2	184	5	US-08-920-440B-10	Sequence 10, Appli
32	176	7.2	184	5	PCT-US95-15781-2	Sequence 2, Appli
33	176	7.2	209	4	US-08-455-001-2	Sequence 7, Appli
34	176	7.2	209	4	US-08-308-814-2	Sequence 2, Appli
35	176	7.2	209	5	PCT-US95-11869-2	Sequence 2, Appli
36	175.5	7.2	213	2	US-09-609-324A-10	Sequence 10, Appli
37	175.5	7.2	213	4	US-08-920-440B-10	Sequence 10, Appli
38	175.5	7.2	213	4	US-09-173-492-10	Sequence 10, Appli
39	175.5	7.1	213	4	US-09-173-133-10	Sequence 10, Appli
40	174.5	7.1	233	4	US-09-214-631-7	Sequence 7, Appli
41	172	7.0	200	1	US-08-455-001-4	Sequence 4, Appli
42	172	7.0	200	5	PCT-US95-11869-4	Sequence 4, Appli
43	169.5	6.9	201	1	US-08-240-124-4	Sequence 4, Appli
44	169.5	6.9	201	1	US-08-453-943-4	Sequence 4, Appli
45	169.5	6.9	201	2	US-09-057-121-4	Sequence 4, Appli
46	169.5	6.9	201	4	US-09-358-734-4	Sequence 4, Appli
47	169	6.9	201	4	US-09-214-631-8	Sequence 8, Appli
48	168.5	6.9	209	4	US-09-214-631-6	Sequence 6, Appli
49	167.5	6.8	228	1	US-08-442-248-4	Sequence 6, Appli
50	167.5	6.8	228	1	US-08-440-815-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-09-214-631-3
; Sequence 3, Application US/09214631
; Patent No. 6413730
; GENERAL INFORMATION:
; APPLICANT: Holland, Sacha
; APPLICANT: Mbamalu, Geraldine
; APPLICANT: Pawson, Tony
; TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED
; TITLE OF INVENTION: BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR
; TITLE OF INVENTION: TYROSINE KINASES
; FILE REFERENCE: 11757.23USWO
; CURRENT APPLICATION NUMBER: US/09/214,631
; CURRENT FILING DATE: 1999-03-12
; EARLIER APPLICATION NUMBER: PCT/CA97/00473
; EARLIER FILING DATE: 1997-07-04
; EARLIER APPLICATION NUMBER: 60/021,272
; EARLIER FILING DATE: 1996-07-05
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 340
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-214-631-3

Query Match 75.0%; Score 1837; DB 4; Length 340;
Best Local Similarity 99.7%;
Matches 337; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy   1 MGPPHSGPGGVRVGALLLLGVLGLVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDL 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MGPPHSGPGGVRVGALLLLGVLGLVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDL 60

Qy  61 LCPRARPGPHSSPNYEFKLYLVLVGGAGQGRRCEAPPAPNLLLTCDRPDLDLRFTIKFQEY 120
         ||||||| |||||||||| |  ||||||||||||||||||||||||||||||||||||||
Db  61 LCPRARPPGPHSSPNYEBYKLYLVVGGAGQGRRCEAPPAPNLLLTCDRPDLDLRFTIKFQEY 120

Qy 121 SPNLWGHEFRSHHDYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKP 180
         |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db 121 SPNLWGHEFRSHHDYYIIATSDGTREGLESLQGGVCLTRGMKVLLQVGQSPRGGAVPRKP 180

Qy 181 VSEMPMEERDRGAAHSLEPGKENLPGDPTPSNATSRGAEGPLPPSMPAVAGAAGGLALLL 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 VSEMPMEERDRGAAHSLEPGKENLPGDPTPSNATSRGAEGPLPPSMPAVAGAAGGLALLL 240
```

```
Qy  241 GVAGAGGANCWRRRAKPSESRHPGPGSFGRGGSLGLGGGGNGPEREAEPGELGIALRGG 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 GVAGAGGANCWRRRAKPSESRHPGPGSFGRGGSLGLGGGGNGPEREAEPGELGIALRGG 300

Qy  301 GAADPPFCPHYEKVSGDYGHPVYIVQDGPPQSPPNIYY 338
        ||||||||||||||||||||||||||||||||||||||
Db  301 GAADPPFCPHYEKVSGDYGHPVYIVQDGPPQSPPNIYY 338
```

RESULT 2
US-08-436-044-2
; Sequence 2, Application US/08436044
; Patent No. 5624899
; GENERAL INFORMATION:
; APPLICANT: Bennett, Brian D.
; APPLICANT: Matthews, William
; TITLE OF INVENTION: HTK LIGAND
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/436,044
; FILING DATE: 05-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/277722
; FILING DATE: 20-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lee, Wendy M.
; REGISTRATION NUMBER: 00,000
; REFERENCE/DOCKET NUMBER: 902D3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1994
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 336 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-436-044-2

Query Match   25.7%; Score 628.5; DB 1; Length 336;
Best Local Similarity 41.8%; Pred. No. 5e-45;
Matches 137; Conservative 49; Mismatches 129; Indels 13; Gaps 5;

```
Qy   14 GALLLLGVLGLVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDLLCPRARPPGPHSS 73
        |:|:|                ||||||:|:|||||||||||||:||::||
Db   17 GLLMVLCRTAISRSIVLEPIYWNSNSKFLPGQGLVLYPQIGDKLDIICPKV---DSKTV 73

Qy   74 PNYEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDLDLRFTIKFQEYSPNLWGHEFRSHH 133
        |:|:|:||:|  ||||||||     |:||||||||:||||||||||||:|:
Db   74 GQYEYKVYVMVDKDQADRCTIKKENTPLLNCARPDQDVKFTIKFQEFSPNLWGLEFQKNK 133

Qy  134 DYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQ--SPRGGAVPRKPVSEMPMER-DR 190
        |||||:|  ||||||:|:||||||:||:|||:||||  |::||  |::|::||
Db  134 DYYIISTSNGSLEGLDNQEGGVCQTRAMKILMKVGQDASSAGSARNHGPTRRPELEAGTN 193

Qy  191 GAAHSLEPGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAGGLALLLLGVGAGGAMC 250
        : : ||||||||   :::||:|::|||:||::|||:
Db  194 GRSSTTSPFVKPNPGSSTDGNSAGHSGNNLLGSEVALFAGIASGCIIPIVIIIITLVVLLL 253
```

```
Qy  251 WRRRAKPSESRHPGPGSFGRGGSLGLGGGGMGPEREAEPGELGIALRGGGAAADPPFCPH 310
        ||||   : :|     ||                    :|  |
Db  254 KYRRRHRKHSPQHTTTLSLSTLIATPKRGGNN----NGSEPSDVIIPLR---TADSVFCPH 306

Qy  311 YEKVSGDYGHPVYIVQDGPPQSPPNIYY 338
        |||  || ||||||||||||:|||:|||
Db  307 YEKVSGDYGHPVYIVQEMPPQSPANIYY 334
```

RESULT 3
US-08-436-054-2
; Sequence 2, Application US/08436054
; Patent No. 5864020
; GENERAL INFORMATION:
; APPLICANT: Bennett, Brian D.
; APPLICANT: Matthews, William
; TITLE OF INVENTION: HTK LIGAND
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/436,054
; FILING DATE: 05-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/277722
; FILING DATE: 20-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lee, Wendy M.
; REGISTRATION NUMBER: 00,000
; REFERENCE/DOCKET NUMBER: 902D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1994
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 336 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-436-054-2

Query Match   25.7%; Score 628.5; DB 2; Length 336;
Best Local Similarity 41.8%; Pred. No. 5e-45;
Matches 137; Conservative 49; Mismatches 129; Indels 13; Gaps 5;

```
Qy   14 GALLLLGVLGLVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDLLCPRARPPGPHSS 73
        |:|:|                ||||||:|:|||||||||||||:||::||
Db   17 GLLMVLCRTAISRSIVLEPIYWNSNSKFLPGQGLVLYPQIGDKLDIICPKV---DSKTV 73

Qy   74 PNYEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDLDLRFTIKFQEYSPNLWGHEFRSHH 133
        |:|:|:||:|  ||||||||     |:||||||||:||||||||||||:|:
Db   74 GQYEYKVYVMVDKDQADRCTIKKENTPLLNCARPDQDVKFTIKFQEFSPNLWGLEFQKNK 133

Qy  134 DYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQ--SPRGGAVPRKPVSEMPMER-DR 190
        |||||:|  ||||||:|:||||||:||:|||:||||  |::||  |::|::||
Db  134 DYYIISTSNGSLEGLDNQEGGVCQTRAMKILMKVGQDASSAGSARNHGPTRRPELEAGTN 193

Qy  191 GAAHSLEPGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAGGLALLLLGVGAGGAMC 250
        : : ||||||||   :::||:|::|||:||::|||:
Db  194 GRSSTTSPFVKPNPGSSTDGNSAGHSGNNLLGSEVALFAGIASGCIIPIVIIIITLVVLLL 253

Qy  251 WRRRRAKPSESRHPGPGSFGRGGSLGLGGGGMGPEREAEPGELGIALRGGGAADPPFCPH 310
```

```
Db  254  KYRRHRKHSPQHTTTLSLSTLATPKRGGNN----NGSEPSDVIIPLR---  |||  |||  TADSVFCPH 306
Qy  311  YEKVSGDYGHPVIVQDGPPQSPPNIYY 338
                                           |:||  ||  |||||
Db  307  YEKVSGDYGHPVIVQEMPPQSPANIYY 334
```

RESULT 4
PCT-US95-08812-2
; Sequence 2, Application PC/TUS9508812
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; TITLE OF INVENTION: HTK LIGAND
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/08812
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Lee, Wendy M.
; REGISTRATION NUMBER: 00,000
; REFERENCE/DOCKET NUMBER: 902PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1994
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 336 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

PCT-US95-08812-2

Query Match   25.7%; Score 628.5; DB 5; Length 336;
Best Local Similarity 41.8%; Pred. No. 5e-45;
Matches 137; Conservative 49; Mismatches 129; Indels 13; Gaps 5;

```
Qy   14  GALILLGVLGLVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGRLDLLCPRARPPGPHSS 73
         |:|: |||||| :|||: |||||||||||||||||:|||:||||
Db   17  GLLMVLCRTAISRSIVLEPIYWNSNSKFLPGGLVLYPQIGDKLDIICPKV---DSKTV 73

Qy   74  PNYEYFKLYLVGGAQGRRCEAPPAPNLLLTCDRPDLDLRFTIKFQEYSPNLWGHEFRSHH 133
         ||||:||||:|||||:|||||||||||||||||:||||||||||||||:
Db   74  GQYEYYKVYMVDKDQADRCTIKKENTPLINCARPDQDVKFTIKFQEBFSPNLWGLEFQKNK 133

Qy  134  DYIILATSDGTREGLESLQGGVCLTRGMKVLLRVGQ--SPRGGAVPRKPVSEMPER-DR 190
         ||:||:||:|||:|||||||||||||||||||||:      ::||||||::
Db  134  DYIILSTSNGSLEGLDNQGGVCQTRAMKILMKVGQDASSAGSARNHGPTRRPELEAGTN 193

Qy  191  GAAHSLEPGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAGGLALLLLGVAGAGGAMC 250
         ||||||:||||||||||||||||||||||||||||||||||:
Db  194  GRSSTTSPFVKPNPGSSTDGNSAGHSGNNLLGSEVALFAGIASGCIIFIVIILVLLL 253

Qy  251  WRRRRAKPSPSESRHPGPGSFQRGGSIGLGLGGGGMGPREAPFGELGIALRGGGAADPFFGPH 310
         |||| :||:||:||||||||||||||
Db  254  KYRRHRKHSPQHTTTLSLSTLATPKRGGNN----NGSEPSDVIIPLR---TADSVFCPH 306

Qy  311  YEKVSGDYGHPVIVQDGPPQSPPNIYY 338
         |||||||||||||||:
Db  307  YEKVSGDYGHPVIVQEMPPQSPANIYY 334
```

RESULT 5
US-08-213-403-2
; Sequence 2, Application US/08213403
; Patent No. 5512457
; GENERAL INFORMATION:
; APPLICANT: Lyman, Stewart D.
; APPLICANT: Beckmann, M. Patricia
; APPLICANT: Baum, Peter R.
; APPLICANT: Carpenter, Melissa
; TITLE OF INVENTION: No. 5512457e1 Cytokine Designated elk ligand
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Immunex Corporation
; STREET: 51 University Street
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Apple System 7.1
; SOFTWARE: Microsoft Word for Apple, Version 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/213,403
; FILING DATE: 15-MAR-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/977,693
; FILING DATE: 13-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Seese, Kathryn A.
; REGISTRATION NUMBER: 32,172
; REFERENCE/DOCKET NUMBER: 2807-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 587-0430
; TELEFAX: (206) 233-0644
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 346 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-213-403-2

Query Match   25.4%; Score 623; DB 1; Length 346;
Best Local Similarity 39.2%; Pred. No. 1.5e-44;
Matches 143; Conservative 48; Mismatches 116; Indels 58; Gaps 9;

```
Qy    8  PGGVRVGALILLGVLGLVSGL-----SLEPVYWNSANKRFQAEGGYVLYPQIGRLDLL 61
         |:|  ||||||||||||||        |||||||||||||||||||||:|||:||||
Db    4  PGQRWLGKWLVAMVVWALCRLATPLAKNLEPVSWSSLNPKFLSGKGLVIYPKIGDKLDII 63

Qy   62  CPRARPPGPHSSPNYEYFKLYLVGGAQGRRCEAPPAPNLLLTCDRPDLDLRFTIKFQEYS 121
         ||||  |   |:|||:||||| |     ||  | :| ||:||| |||:|||||||||:
Db   64  CPRAEAGRP-----YEYYKLYLVRPEQAAACSTVLDPNVLVTCNRPEQEIRFTIKFQEFS 118

Qy  122  PNLWGHEFRSHHDYIILATSDGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKPV 181
         ||:||:|| |:|||| :|:|:|:|||| |||||| ||:||:::|:|||
Db  119  PNIWGLEFKKHHDYIYTSTSNGSLEGLENREGGVCRTRTMKIIMKVGQDPNAVTPEQLTT 178

Qy  182  SEMPMERDRGAAHSLE-PGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAGGLA---- 236
         |:        | |    ||  :|| |:| | :: ||  | |  |||||||||||:
Db  179  SRPSKEADNTVKMATQAPGSRGSLGDSDGKHETVNQEEKSGP------GASGGSSGDPD 231

Qy  237  -----LILLGVAGAGGA-----------MCWRRRRAKPSESRHPGPGSFQRGGGSLGL 277
              |:||||| || |           |:| |||||:|||:|||||||
Db  232  GFFNSKVALFAAVGAGCVIFLLIIFITVLLLKLRKRHRKHTQQ------RAAALSL 282
```

RESULT 6
US-08-458-077-2
; Sequence 2, Application US/08458077
; Patent No. 5627267
; GENERAL INFORMATION:
; APPLICANT: Lyman, Stewart D.
; APPLICANT: Beckmann, M. Patricia
; APPLICANT: Baum, Peter R
; APPLICANT: Carpenter, Melissa
; TITLE OF INVENTION: No. 5627267e1 Cytokine Designated elk Ligand
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Immunex Corporation
; STREET: 51 University Street
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Apple System 7.1
; SOFTWARE: Microsoft Word for Apple, Version 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/458,077
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/213,403
; FILING DATE: 15-MAR-1994
; APPLICATION NUMBER: US 07/977,693
; FILING DATE: 13-NOV-1992
; CLASSIFICATION: 514
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 346 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-458-077-2

```
Query Match            25.4%;   Score 623;  DB 1;  Length 346;
Best Local Similarity  39.2%;   Pred. No. 1.5e-44;
Matches  143;   Conservative  48;   Mismatches 116;   Indels  58;   Gaps  9;

Qy      8  PGGVRVGALLLLGVGLVSGL-------SLEPVYWNSANKRFQAEGGYVLYPQIGDRLDLL  61
           ||  | |||| ||| || ||       |||| |||||||||||| ||||||||||||||
Db      4  PGQRWLGKWLVAMVWALCRLATPLAKNLEPVSWSSLNPKFLSGKGLVIYPKIGDKLDII  63

Qy     62  CPRARPPPGPHSSNYEFYKYLYLYGGAQGRRCREAPPAPNLLLTCDRPDLDLRFTIKFQEYS  121
           | |  |  || |  || ||| || |||| | |  |  |  ||||||||||||||| |||||
Db     64  CPRAEAGRP----YEYYKLYLVRPEQAAACSTVLDPNVLVTICNRPEQEIRFTIKFQEFS  118

Qy    122  PNLWGHEFRSHHDYYIIATSDGTREGLESLQGGVCLTRGMKYLLRVGQSPRGGAVPRKPV  181
           ||| | |||| || |||||  | ||||  |  |||| |||||||||||  ||   ||| |
Db    119  PNTWGLEFKHHDYIITSTNGSLEGLENREGGVCRTRTMKIIMKVGQDPNAVTPEQLTT  178
```

RESULT 7
US-08-460-741-2
; Sequence 2, Application US/08460741
; Patent No. 5670625
; GENERAL INFORMATION:
; APPLICANT: Lyman, Stewart D.
; APPLICANT: Beckmann, M. Patricia
; APPLICANT: Baum, Peter R
; APPLICANT: Carpenter, Melissa
; TITLE OF INVENTION: No. 5670625e1 Cytokine Designated elk Ligand
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Immunex Corporation
; STREET: 51 University Street
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Apple System 7.1
; SOFTWARE: Microsoft Word for Apple, Version 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,741
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/213,403
; FILING DATE: 15-MAR-1994
; APPLICATION NUMBER: US 07/977,693
; FILING DATE: 13-NOV-1992
; CLASSIFICATION: 530
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 346 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-460-741-2

```
Query Match            25.4%;   Score 623;  DB 1;  Length 346;
Best Local Similarity  39.2%;   Pred. No. 1.5e-44;
Matches  143;   Conservative  48;   Mismatches 116;   Indels  58;   Gaps  9;

Qy      8  PGGVRVGALLLLGVGLVSGL-------SLEPVYWNSANKRFQAEGGYVLYPQIGDRLDLL  61
           ||  | |||| ||| || ||       |||| |||||||||||| ||||||||||||||
Db      4  PGQRWLGKWLVAMVWALCRLATPLAKNLEPVSWSSLNPKFLSGKGLVIYPKIGDKLDII  63

Qy     62  CPRARPPPGPHSSNYEFYKYLYLYGGAQGRRCREAPPAPNLLLTCDRPDLDLRFTIKFQEYS  121
           | |  |  || |  || ||| || |||| | |  |  |  ||||||||||||||| |||||
Db     64  CPRAEAGRP----YEYYKLYLVRPEQAAACSTVLDPNVLVTICNRPEQEIRFTIKFQEFS  118

Qy    122  PNLWGHEFRSHHDYYIIATSDGTREGLESLQGGVCLTRGMKYLLRVGQSPRGGAVPRKPV  181
           ||| | |||| || |||||  | ||||  |  |||| |||||||||||  ||   ||| |
Db    119  PNTWGLEFKHHDYIITSTNGSLEGLENREGGVCRTRTMKIIMKVGQDPNAVTPEQLTT  178

Qy    182  SEMPMERDRGRAHSLE-PGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAGGLA----  236
           |  | | | || ||  ||| ||||| ||| |||||| || ||||| |||| || |
Db    179  SRPSKEADNTVRMATCQAPGRGSLGDSDGKHETVNQBEKSGP------GASGGSSGGDPD  231

Qy    237  -----LLLGVAGAGGA-------------MCWRRRAKPSESRHPGGSSFGRGGSLGL  277
                |||| ||| ||               | |   |  |     ||   | |||
Db    232  GFFNSKVALFAAVGAGCVIFLLIIFLTVLLLKLRRHRKHTQQ----RAAALSL  282

Qy    278  ----GGGGGMGPREAEPGELGIALRGGGAADPPPFCPHYEKVSGDYGHPVYIVQDGPPQSP  333
           ||||||||| |||||||||||||||||| |||| ||||||| |||| ||||||||||| |
Db    283  STLASPKGGSGTAGTEPSDIIIPLR---TTENNYCPHYEKVSGDYGHPVYIVQEMPPQSP  339

Qy    334  PNIYY  338
           |||||
Db    340  ANIYY  344
```

RESULT 8
US-08-747-240-2
; Sequence 2, Application US/08747240
; Patent No. 5728813
; GENERAL INFORMATION:
; APPLICANT: Lyman, Stewart D.
; APPLICANT: Beckmann, M. Patricia
; APPLICANT: Baum, Peter R.
; APPLICANT: Carpenter, Melissa
; TITLE OF INVENTION: No. 5728813e1 Cytokine Designated elk Ligand
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Immunex Corporation
; STREET: 51 University Street
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Apple System 7.1
; SOFTWARE: Microsoft Word for Apple, Version 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/747,240
; FILING DATE: 12-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/213,403
; FILING DATE: 15-MAR-1994
; APPLICATION NUMBER: US 07/977,693
; FILING DATE: 13-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Seese, Kathryn A.
; REGISTRATION NUMBER: 32,172
; REFERENCE/DOCKET NUMBER: 2807-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 587-0430
; TELEFAX: (206) 233-0644
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 346 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-747-240-2

```
Qy   62  CPRARPPGPHSSPNYEFYKLYLVGGAQGRRCEAPAPNLLLTCDRPDLDLRFTIKFQEYS  121
             ||:|||||
Db   64  CPRAEAGRP-----YEYYKLYLVRPEQAAACSTVLDPNVLVTCNRPEQEIRFTIKFQEFS  118

Qy  122  PNLWGHEFRSHHDYIYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKPV  181
             |:|||:||:||||
Db  119  PNYMGLEFPKKHHDYIYITSTSNGSLEGLENRGGVCRTRTMKIIMKVGQDPNAVTPEQLTT  178

Qy  182  SEMPMERDRGAAHSLE-PGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAGGLA----  236
             ||:||
Db  179  SRPSKEADNTVRMATQAPGSRGSLGDSDGKHETVNQEEKSGP-------GASGGSSGDPD  231

Qy  237  -------LLLLGVAGGGA------------MCWRRRAKPSESRHPGPGSFGRGGSLGL  277
             |:|||
Db  232  GFFNSKVALFAAVGAGCVIFLLIIFLTVLLLKLRKRHRKHTQQ-----------RAAALSL  282

Qy  278  ----GGGGGMGPREAEPGELGIALRGGGAADPPFCPHYEKVSGDYGHPVYIVQDGPPQSP  333
             ||
Db  283  STLASPKGGSGTAGTEPSDIIIPLR---TTENNYCPHYEKVSGDYGHPVYIVQEMPPQSP  339

Qy  334  PNIYY  338
         |||||
Db  340  ANIYY  344
```

RESULT 9
US-08-299-567-6
; Sequence 6, Application US/08299567
; Patent No. 5747033
; GENERAL INFORMATION:
; APPLICANT: Davis, et al.
; TITLE OF INVENTION: METHOD OF ENHANCING THE BIOLOGICAL
; TITLE OF INVENTION: ACTIVITY OF EPH FAMILY LIGANDS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Regeneron Pharmaceuticals, Inc.
; STREET: 777 Old Saw Mill River Road
; CITY: Tarrytown
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10591-6707
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/299,567
; FILING DATE: 01-SEP-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kempler, Gail M.
; REGISTRATION NUMBER: 32,143
; REFERENCE/DOCKET NUMBER: REG 290
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 914-345-7400
; TELEFAX: 914-345-7721
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 346 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein

Query Match   25.4%; Score 623; DB 1; Length 346;
Best Local Similarity 39.2%; Pred. No. 1.5e-44;
Matches 143; Conservative 48; Mismatches 116; Indels 58; Gaps 9;

```
Qy    8  PGGVRVGALLLGVLGLIVSGL--------SLEPYYWNSANKRFQAEGGYYLYPQIGDRLDLL  61
            ||:|||||||||||||
Db    4  PGQRWLGKWLVAMVVWWALCRLAATPLAKNLEPVSWSSLNPKFLSGKGLVIYFKIGDKLDII  63

Qy   62  CPRARPPGPHSSPNYEFYKLYLVGGAQGRRCEAPAPNLLLTCDRPDLDLRFTIKFQEYS  121
             ||:|||||
Db   64  CPRAEAGRP-----YEYYKLYLVRPEQAAACSTVLDPNVLVTCNRPEQEIRFTIKFQEFS  118

Qy  122  PNLWGHEFRSHHDYIYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKPV  181
             |:|||:||:||||
Db  119  PNYMGLEFPKKHHDYIYITSTSNGSLEGLENRGGVCRTRTMKIIMKVGQDPNAVTPEQLTT  178

Qy  182  SEMPMERDRGAAHSLE-PGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAGGLA----  236
             ||:||
Db  179  SRPSKEADNTVRMATQAPGSRGSLGDSDGKHETVNQEEKSGP-------GASGGSSGDPD  231

Qy  237  -------LLLLGVAGGGA------------MCWRRRAKPSESRHPGPGSFGRGGSLGL  277
             |:|||
Db  232  GFFNSKVALFAAVGAGCVIFLLIIFLTVLLLKLRKRHRKHTQQ-----------RAAALSL  282

Qy  278  ----GGGGGMGPREAEPGELGIALRGGGAADPPFCPHYEKVSGDYGHPVYIVQDGPPQSP  333
             ||
Db  283  STLASPKGGSGTAGTEPSDIIIPLR---TTENNYCPHYEKVSGDYGHPVYIVQEMPPQSP  339

Qy  334  PNIYY  338
         |||||
Db  340  ANIYY  344
```

US-08-299-567-6

Query Match   25.4%; Score 623; DB 1; Length 346;
Best Local Similarity   39.2%; Pred. No. 1.5e-44;
Matches 143; Conservative 48; Mismatches 116; Indels 58; Gaps 9;

```
Qy    8  PGGVRVGALLILGVILGLIVSGLI------SLEPVYWNSANKRFQAEGGYVLYPQIGDRLDLL   61
Db    4  PGQRWLGKWLVAMVWVWALCRLATPLAKNLEPVSWSSLNPKFLSGKGLVIYPKIGDKLDII    63

Qy   62  CPRARPPGPHSSPNYEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPLDLRFTIKFQEYS     121
Db   64  CPRAEAGRP-----YEYYKLYLVRPEQAAACSTVLDPNVLVTCNRPEQEIRFTIKFQEFS    118

Qy  122  PNLWGHEFRSHHDYYIIATSDGTREGLESSLQGGVCLTRGMKVLLRVQQSPRGGAVPRKPV   181
Db  119  PNYMGLEFKKHHDYYITSTSNGSLEGLENREGGVCRTRTMKIIMKVGQDPNAVTPEQLTT   178

Qy  182  SEMPMERDRGAAHSLE-PGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAGGLA----   236
Db  179  SRPSKEADNTVKMATQAFQSRGSLGDSDGKHETVNQEEKSGP------GASGGSSGDPD   231

Qy  237  -----LILLGVAGAGGA------------MCWRRRAKPSESRHPGPGSFGRGGSLGL     277
Db  232  GFFNSKVALFAAVGAGCVIFLLIIIFLVLLKLRKIRKHTQQ----------RAAALSL   282

Qy  278  ----GGGGGGMGPREAEFGELGIAILRGGGAADPPFCPHYEKVSGDYGHPVYIVQDGPPQSP   333
Db  283  STLASPKGGSGTAGTEPSDIIIPLR---TTENNYCPHYEKVSGDYGHPVYIVQEMPPQSP   339

Qy  334  PNIYY   338
Db  340  ANIYY   344
```

RESULT 10
US-08-436-044-4
; Sequence 4, Application US/08436044
; Patent No. 5624899
; GENERAL INFORMATION:
; APPLICANT: Bennett, Brian D.
; APPLICANT: Matthews, William
; TITLE OF INVENTION: HTK LIGAND
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/436,044
; FILING DATE: 05-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/277722
; FILING DATE: 20-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lee, Wendy M.
; REGISTRATION NUMBER: 00,000
; REFERENCE/DOCKET NUMBER: 902D3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1994
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:

; LENGTH: 333 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

US-08-436-044-4

Query Match   25.3%; Score 620.5; DB 1; Length 333;
Best Local Similarity   40.5%; Pred. No. 2.3e-44;
Matches 133; Conservative 52; Mismatches 130; Indels 13; Gaps 5;

```
Qy   14  GALILLGVILGVILVSGLISLEPVYVWNSANKRFQAEGGYVLYPQIGDRLDLLCPRARPPGPHSS   73
Db   14  GVLIMVLCRTAISKSIVLEPIYVWNSSNSKFLPGGQGLVIYPQIGKLDIIICPKV---DSKTV    70

Qy   74  PNYEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPLDLRFTIKFQEYSPNLWGHEFRSHH    133
Db   71  GQYEYVKVYMWVDKDQADRCTIKKENTPLLNCAKPDQDIKFTIKFQEFSPNLWGLITFQKNK   130

Qy  134  DYYIIATSDGTREGLESSLQGGVCLTRGMKVLLRVQQ--SPRGGAVPRKPVSEMPMER-DR    190
Db  131  DYYIISTSNGSLEGLDNQEGGVCQTRAMKILMKVGQDASSAGSTRNKDPTRRPELEAGTN   190

Qy  191  GAAHSLEPGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAGGLALLLLGVAGAGGAMC   250
Db  191  GRSSTTSPFVXPNPGSSTDGNSAGHSGNNILGSEVALFAGIASGCIIFIVIIILVVLLL   250

Qy  251  WRRRRAKPSESRHPGPGSFGRGGSLGLGLGGGGGGMGPREAEGGELGIALRGGGAADPPFCPH   310
Db  251  KLRRHRHKHSPQHTTTLSLSTLATPKRSGNN----NGSEPSDIIIPLR---TADSVPCPH   303

Qy  311  YEKVSGDYGHPVYIVQDGPPQSPSPPNIYY   338
Db  304  YEKVSGDYGHPVYIVQEMPPQSPSPANIYY   331
```

RESULT 11
US-08-436-054-4
; Sequence 4, Application US/08436054
; Patent No. 5864020
; GENERAL INFORMATION:
; APPLICANT: Bennett, Brian D.
; APPLICANT: Matthews, William
; TITLE OF INVENTION: HTK LIGAND
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/436,054
; FILING DATE: 05-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/277722
; FILING DATE: 20-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lee, Wendy M.
; REGISTRATION NUMBER: 00,000
; REFERENCE/DOCKET NUMBER: 902D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1994
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 333 amino acids

```
;
;  TYPE: amino acid
;  TOPOLOGY: linear
US-08-436-054-4

Query Match     25.3%;  Score 620.5;  DB 2;  Length 333;
Best Local Similarity   40.5%;  Pred. No. 2.3e-44;
Matches 133;  Conservative  52;  Mismatches 130;  Indels  13;  Gaps   5;

Qy    14 GALILLGVLGLVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDLLCPRARPPGHSS 73
         :||:|  ||||| |||||:|||:|||||||:||:|||||||:|
Db    14 GVLMVLCRTAISKSIVLEPIYWNSSNSKFLPGQGLVLYPQIGDKLDIICPKV---DSKTV 70

Qy    74 PNYEFYKLIVLGGAQGRRCEAPPAPNLLLTCDRPDLDLRFTIKFQEYSPNLWGHEFRSHH 133
         ::::|| ::: ||| |: ::|| ||:::| |:||||:::|:|:||:||| |:|||||
Db    71 GQYEYKVYMVDKDQADRCTIKKENTPLINCAKPDQDIKFTIKFQEFSPNLWGLEFQKNK 130

Qy   134 DYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQ--SPRGGAVPRKPVSEMPMER-DR 190
         |::|::|  || |:||| :||||| |||:|||| :|    |:| |:| ||||:|:::
Db   131 DYYIISTNGSLEGLDNQEGGVCQTRAMKILMKVGQDASSAGSTRNKDFTRRPELEAGTN 190

Qy   191 GAAHSLEPGKENLPGDPTSNATSRGAEGPLPPPSMPAVGAAAGGLALLLLGVAGAGGAMC 250
         |:  :  : |  ||| :  ||:  | | |::| :  ||  |::| |||:|||:| ||||
Db   191 GRSSTTSPFVKPNPGSSTDGNSAGHSGNNILGSEVALFAGIASGCIIFIVIIITLIVLLL 250

Qy   251 WRRRRAKPSESRHPGPGSFGRGGSLGLGGGGGMGPREAEPGELGIALRGGGAADPPFCPH 310
         : :    :: |:|  | :|:|||:|  ::: |:|  |||||: |||  ::  |  | ||
Db   251 KYRRRHRKHSPQHTTTLSLSTLATPKRSGNN----NGSEPSDIIPLR----TADSVFCPH 303

Qy   311 YEKVSGDYGHPVYIVQDGPPQSPPNIYY 338
         ||||||||||||:||||||||||:| ||||
Db   304 YEKVSGDYGHPVIVQEMPPQSPANIYY 331


RESULT 12
US-08-271-948-2
Sequence 2, Application US/08271948
Patent No. 6303769
;  GENERAL INFORMATION:
;  APPLICANT: Cerretti, Douglas P.
;  APPLICANT: Reddy, Pranitha
;  TITLE OF INVENTION: No. 630376951 Cytokine Designated lerk-5
;  NUMBER OF SEQUENCES: 3
;  CORRESPONDENCE ADDRESS:
;  ADDRESSEE: Immunex Corporation
;  STREET: 51 University Street
;  CITY: Seattle
;  STATE: Washington
;  COUNTRY: US
;  ZIP: 98101
;  COMPUTER READABLE FORM:
;  MEDIUM TYPE: Floppy disk
;  COMPUTER: Apple Macintosh
;  OPERATING SYSTEM: Apple 7.1
;  SOFTWARE: Microsoft Word, Version 5.1a
;  CURRENT APPLICATION DATA:
;  APPLICATION NUMBER: US/08/271,948
;  FILING DATE:
;  CLASSIFICATION: 435
;  ATTORNEY/AGENT INFORMATION:
;  NAME: Seese, Kathryn A.
;  REGISTRATION NUMBER: 32,172
;  REFERENCE/DOCKET NUMBER: 2823
;  TELECOMMUNICATION INFORMATION:
;  TELEPHONE: (206) 587-0430
;  TELEFAX: (206) 233-0644
;  TELEX: 756822
;  INFORMATION FOR SEQ ID NO: 2:
;  SEQUENCE CHARACTERISTICS:
;  LENGTH: 333 amino acids
;  TYPE: amino acid
;  TOPOLOGY: linear
;  MOLECULE TYPE: protein
US-08-271-948-2
```

```
RESULT 13
PCT-US95-08534-2
;  Sequence 2, Application PC/TUS950853534
;  GENERAL INFORMATION:
;  APPLICANT: Immunex Corporation
;  TITLE OF INVENTION: Novel Cytokine Designated lerk-5
;  NUMBER OF SEQUENCES: 3
;  CORRESPONDENCE ADDRESS:
;  ADDRESSEE: Immunex Corporation
;  STREET: 51 University Street
;  CITY: Seattle
;  STATE: Washington
;  COUNTRY: US
;  ZIP: 98101
;  COMPUTER READABLE FORM:
;  MEDIUM TYPE: Floppy disk
;  COMPUTER: Apple Macintosh
;  OPERATING SYSTEM: Apple 7.1
;  SOFTWARE: Microsoft Word, Version 5.1a
;  CURRENT APPLICATION DATA:
;  APPLICATION NUMBER: PCT/US95/08534
;  FILING DATE: 06-JUL-1995
;  CLASSIFICATION:
;  PRIOR APPLICATION DATA:
;  APPLICATION NUMBER: US 08/271,948
;  FILING DATE: 08-JUL-1994
;  CLASSIFICATION:
;  ATTORNEY/AGENT INFORMATION:
;  NAME: Anderson, Kathryn A.
;  REGISTRATION NUMBER: 32,172
;  REFERENCE/DOCKET NUMBER: 2823-WO
;  TELECOMMUNICATION INFORMATION:
;  TELEPHONE: (206) 587-0430
;  TELEFAX: (206) 233-0644
;  TELEX: 756822
;  INFORMATION FOR SEQ ID NO: 2:
;  SEQUENCE CHARACTERISTICS:
;  LENGTH: 333 amino acids
;  TYPE: amino acid
;  TOPOLOGY: linear
;  MOLECULE TYPE: protein
PCT-US95-08534-2

Query Match     25.3%;  Score 620.5;  DB 4;  Length 333;
Best Local Similarity   40.5%;  Pred. No. 2.3e-44;
Matches 133;  Conservative  52;  Mismatches 130;  Indels  13;  Gaps   5;

Qy    14 GALILLGVLGLVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDLLCPRARPPGHSS 73
         :||:|  ||||| |||||:|||:|||||||:||:|||||||:||
Db    14 GVLMVLCRTAISKSIVLEPIYWNSSNSKFLPGQGLVLYPQIGDKLDIICPKV---DSKTV 70

Qy    74 PNYEFYKLIVLGGAQGRRCEAPPAPNLLLTCDRPDLDLRFTIKFQEYSPNLWGHEFRSHH 133
         ::::|| ::: ||| |: ::|| ||:::| |:||||:::|:|:||:||| |:|||||
Db    71 GQYEYKVYMVDKDQADRCTIKKENTPLINCAKPDQDIKFTIKFQEFSPNLWGLEFQKNK 130

Qy   134 DYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQ--SPRGGAVPRKPVSEMPMER-DR 190
         |::|::|  || |:||| :||||| |||:|||| :|    |:| |:| ||||:|:::
Db   131 DYYIISTNGSLEGLDNQEGGVCQTRAMKILMKVGQDASSAGSTRNKDFTRRPELEAGTN 190

Qy   191 GAAHSLEPGKENLPGDPTSNATSRGAEGPLPPPSMPAVGAAAGGLALLLLGVAGAGGAMC 250
         |:  :  : |  ||| :  ||:  | | |::| :  ||  |::| |||:|||:| ||||
Db   191 GRSSTTSPFVKPNPGSSTDGNSAGHSGNNILGSEVALFAGIASGCIIFIVIIITLIVLLL 250

Qy   251 WRRRRAKPSESRHPGPGSFGRGGSLGLGGGGGMGPREAEPGELGIALRGGGAADPPFCPH 310
         : :    :: |:|  | :|:|||:|  ::: |:|  |||||: |||  ::  |  | ||
Db   251 KYRRRHRKHSPQHTTTLSLSTLATPKRSGNN----NGSEPSDIIPLR----TADSVFCPH 303

Qy   311 YEKVSGDYGHPVYIVQDGPPQSPPNIYY 338
         ||||||||||||:||||||||||:| ||||
Db   304 YEKVSGDYGHPVIVQEMPPQSPANIYY 331
```

RESULT 14
PCT-US95-08812-4
; Sequence 4, Application PC/TUS9508812
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; TITLE OF INVENTION: HTK LIGAND
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/08812
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Lee, Wendy M.
; REGISTRATION NUMBER: 00,000
; REFERENCE/DOCKET NUMBER: 902PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1994
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 333 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
PCT-US95-08812-4

  Query Match   25.3%; Score 620.5; DB 5; Length 333;
  Best Local Similarity   40.5%; Pred. No. 2.3e-44;
  Matches   133; Conservative   52; Mismatches 130; Indels 13; Gaps   5;

Qy   14 GAILLLGVILGLVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDLLCPRARPPGPHSS 73
              |::|.| | .||||||:||||:|||||||:|||.|||||||.|||||||||
Db   14 GVLMVLCRTAISKSIVLEPIYWNSNSKFLPGQGLVLYPQIGDKLDIICPKV---DSKTV 70

Qy   74 PNYEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDLDLRFTIKFQEYSPNLWGHEFRSHH 133
             ||||| |||:|||||||:|| :|::||||.|:|||||:||:|||||:||||||||:|||
Db   71 GQYEYKVYMVDKDQADRCTIKKENTPLLNCAKPDQDIKFTIKFQEFSPNLWGLEFQKNK 130

Qy   134 DYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQ--SPRGGAVPRKPVSEMPER-DR 190
              |:||.|||||:|||.||.||||||||||||::||: .|::|||:|||||||:|.||
Db   131 DYYIISTSNGSLEGLDNQEGGVCQTRAMKILMKVGQDASSAGSTRNKDPTRRPELEAGTN 190

Qy   191 GAAHSLEPGKENLPGDPISNATSRGAEGPLPPPSMPAVAGAAGGLALLLLGVAGAGGAMC 250
             |||||||||||||||||.|:||||||||||||.|||||||||||:|||||||||||||||
Db   191 GRSSTTSPFVKPNPGSSTDGNSAGHSGNNILGSEVALFAGIASGCIFIVIIITLVVLLL 250

Qy   251 WRRRRAKPSESRHPGPGSFGRGGSLGLGGGGGMGPREAEPGELGIALRGGGAADPPFCPH 310
             :|:|::|::::||.||:|:|||||||||||||:||:||:|||||:|:||||||:|||||
Db   251 KYRRHRKHSPQHTTLSLSTLATPKRSGNN----NGSEPSDIIIPLR---TADSVFCPH 303

Qy   311 YEKVSGDYGHPVYIVQDGPPQSPPNIYY 338
             ||||||||||||||||||||||||||||
Db   304 YEKVSGDYGHPVYIVQEMPPQSPANIYY 331

RESULT 15
US-09-214-631-4
; Sequence 4, Application US/09214631
; Patent No. 6413730
; GENERAL INFORMATION:
; APPLICANT: Holland, Sacha
; APPLICANT: Mbamalu, Geraldine
; APPLICANT: Pawson, Tony
; TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED
; TITLE OF INVENTION: BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR
; TITLE OF INVENTION: TYROSINE KINASES
; FILE REFERENCE: 11757.23USWO
; CURRENT APPLICATION NUMBER: US/09/214,631
; CURRENT FILING DATE: 1999-03-12
; EARLIER APPLICATION NUMBER: US/09/214,631
; EARLIER FILING DATE: 1999-07-05
; EARLIER APPLICATION NUMBER: PCT/CA97/00473
; EARLIER FILING DATE: 1997-07-04
; EARLIER APPLICATION NUMBER: 60/021,272
; EARLIER FILING DATE: 1996-07-15
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-214-631-4

  Query Match   25.0%; Score 613.5; DB 4; Length 333;
  Best Local Similarity   40.5%; Pred. No. 9e-44;
  Matches   133; Conservative   54; Mismatches 128; Indels 13; Gaps   6;

Qy   14 GAILLLGVILGLVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDLLCPRARPPGPHSS 73
              |::|.| | .||||||:||||:|||||||:|||.|||||||.|||||||||
Db   14 GVLMVLCRTAISKSIVLEPIYWNSNSKFLPGQGLVLYPQIGDKLDIICPKV---DSKTV 70

Qy   74 PNYEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDLDLRFTIKFQEYSPNLWGHEFRSHH 133
             ||||| |||:|||||||:|| :|::||||.|:|||||:||:|||||:||||||||:|||
Db   71 GQYEYKVYMVDKDQADRCTIKKENTPLLNCAKPDQDIKFTIKFQEFSPNLWGLEFQKNK 130

Qy   134 DYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQ--SPRGGAVPRKPVSEMP-MER-DR 190
              |:||.|||||:|||.||.||||||||||||::||: .|::|||:|||||||:|.||
Db   131 DYYIISTSNGSLEGLDNQEGGVCQTRAMKILMKVGQDASSAGSTRNKDPTRRPELEAGTN 190

Qy   191 GAAHSLEPGKENLPGDPISNATSRGAEGPLPPPSMPAVAGAAGGLALLLLGVAGAGGAMC 250
             |||||||||||||||||.|:||||||||||||.|||||||||||:|||||||||||||||
Db   191 GRSSTTSPFVKPNPGSSTDGNSAGHSGNNILGSEVALFAGIASGCIFIVIIITLVVLLL 250

Qy   251 WRRRRAKPSESRHPGPGSFGRGGSLGLGGGGGMGPREAEPGELGIALRGGGAADPPFCPH 310

```
Db    251 KYRRHRKHSPQHTTTLSLSTLIATPKRSGNN----NGSEPSDIIIPLR---TADSVFPH 303
              |||               |::  |||         |:|:|          ||   |||
Qy    311 YEKVSGDYGHPVYIVQDGPPQSPPNIIY 338
              |||||||||||||||  |||||||||
Db    304 YEKVSGDYGHPVYIVQEMPPQSPPNIIY 331
```

; APPLICANT: Mbamalu, Geraldine
; APPLICANT: Pawson, Tony
; TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED
; TITLE OF INVENTION: BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR
; TITLE OF INVENTION: TYROSINE KINASES
; FILE REFERENCE: 11757.23USWO
; CURRENT APPLICATION NUMBER: US/09/214,631
; CURRENT FILING DATE: 1999-03-12
; EARLIER APPLICATION NUMBER: PCT/CA97/00473
; EARLIER FILING DATE: 1997-07-04
; EARLIER APPLICATION NUMBER: 60/021,272
; EARLIER FILING DATE: 1996-07-05
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 13
; LENGTH: 89
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-214-631-13

Query Match   100.0%; Score 489; DB 4; Length 89;
Best Local Similarity   100.0%; Pred. No. 4.5e-34;
Matches   87; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy    252 RRRRAKPSESRHPGPSFGRGGSLGLGGGGGWGPREAEPGELGIALRGGGAADPPFGPHY 311
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 RRRRAKPSESRHPGPSFGRGGSLGLGGGGGWGPREAEPGELGIALRGGGAADPPFGPHY 60

Qy    312 EKVSGDYGHPVYIVQDGPPQSPPNIYY 338
              ||||||||||||||||||||||||||
Db     61 EKVSGDYGHPVYIVQDGPPQSPPNIYY 87
```

RESULT 18
US-09-214-631-11
; Sequence 11, Application US/09214631
; Patent No. 6413730
; GENERAL INFORMATION:
; APPLICANT: Holland, Sacha
; APPLICANT: Mbamalu, Geraldine
; APPLICANT: Pawson, Tony
; TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED
; TITLE OF INVENTION: BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR
; TITLE OF INVENTION: TYROSINE KINASES
; FILE REFERENCE: 11757.23USWO
; CURRENT APPLICATION NUMBER: US/09/214,631
; CURRENT FILING DATE: 1999-03-12
; EARLIER APPLICATION NUMBER: PCT/CA97/00473
; EARLIER FILING DATE: 1997-07-04
; EARLIER APPLICATION NUMBER: 60/021,272
; EARLIER FILING DATE: 1996-07-05
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 11
; LENGTH: 82
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-214-631-11

Query Match   20.0%; Score 192; DB 4; Length 82;
Best Local Similarity   46.2%; Pred. No. 3.2e-09;
Matches   42; Conservative   9; Mismatches   24; Indels   16; Gaps   3;

```
Qy    252 RRRRAKPSESRHPGPSFGRGGSLGL----GGGGGWGPREAEPGELGIALRGGGAADPPFF 307
              || :||     |||||||||||||     |:|  | ||      :|    |        |
Db      2 RKRHRKHTQQ----------RAAALSLTLASPKGGSGTAGTEPSDIIIPLR---TTENNY 49

Qy    308 CPHYEKVSGDYGHPVYIVQDGPPQSPPNIYY 338
              |||:|::||||||||||||||||||||
Db     50 CPHYEKVSGDYGHPVYIVQEMPPQSPPNIYY 80
```

RESULT 19

RESULT 16
US-09-214-631-5
; Sequence 5, Application US/09214631
; Patent No. 6413730
; GENERAL INFORMATION:
; APPLICANT: Holland, Sacha
; APPLICANT: Mbamalu, Geraldine
; APPLICANT: Pawson, Tony
; TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED
; TITLE OF INVENTION: BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR
; TITLE OF INVENTION: TYROSINE KINASES
; FILE REFERENCE: 11757.23USWO
; CURRENT APPLICATION NUMBER: US/09/214,631
; CURRENT FILING DATE: 1999-03-12
; EARLIER APPLICATION NUMBER: PCT/CA97/00473
; EARLIER FILING DATE: 1997-07-04
; EARLIER APPLICATION NUMBER: 60/021,272
; EARLIER FILING DATE: 1996-07-05
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 5
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-214-631-5

Query Match   24.5%; Score 599.5; DB 4; Length 345;
Best Local Similarity   38.4%; Pred. No. 1.4e-42;
Matches 140; Conservative 50; Mismatches 116; Indels 59; Gaps 10;

```
Qy      8 PGGVRVG----ALLLLGVLGLVSGL--SLEPVTYWNSANKFQAEGGYVLYPQIGDRLDLL 61
              |||         |:|    |:||    ||:    :|  |         ::   | ||||
Db      4 PGGRWLGKWLYAMVVWALCRLATPLAKNLEPVSWSSLNPKFLSGKGLVIYPKIGDKLDIi 63

Qy     62 CPRARPPGPHSSPNYEFYKLYLVGAQGRRCEAPPAPNLLLTCRPDLLRFTIKFQEYS 121
              ||          ::|||||||  |:  ||::    ::|||::|||:|||||||||:
Db     64 CPPEAGRP-----YEYYKLYLVRPEQAAACSTVLDPMVLVTCNRPEQEIRFTIKFQEFS 118

Qy    122 PNLWGHEFRSHHDYIYIIATSDGTREGLESLQGGYCLTRGMKVLLRVGQSPRGGAVPRKPV 181
              ||  | ||: ||||||  ::| : ||||:|::|| | |:|||:: |||: |  | |::
Db    119 PNYMGLEFKKHHDYIYTSTSNGSLEGLENREGGVCRTRTMKIIKVGQDPNAVTPEQLTT 178

Qy    182 SEMPMERDRGAAHSLE-PGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAGGLA---- 236
              |            |  | : |  ||  |: |  |::  |           |  | |
Db    179 SRPSKEADNTVKMATQAPGSRGSLGDSDGKHEETVNQEEKSGP------GASGSSSGDD 231

Qy    237 -----LLLLGVAGAGGA------MCWRRRRAKPSESRHPGPSFGRGGSLGL 277
                   ||| ||||| |       ||:||||||||||||||||||||||||
Db    232 GFFNSKVALFAAVGAGCVIFLLLIIIFLTVLLLKLPKRHRKHTQ------RAAALSL 281

Qy    278 ----GGGGGWGPREAEPGELGIALRGGGAADPPFGPHYEKVSGDYGHPVYIVQDGPPQSP 333
                  |||||||||||||||||||||||||||  |||||||||||||||||||  ||||||
Db    282 STIASPKGGSGTAGTEPSDIIIPL---FTTENNYCPHYEKVSGDYGHPVYIVQEMPPQSP 338

Qy    334 PNIYY 338
              ||||
Db    339 ANIYY 343
```

RESULT 17
US-09-214-631-13
; Sequence 13, Application US/09214631
; Patent No. 6413730
; GENERAL INFORMATION:
; APPLICANT: Holland, Sacha

```
; US-09-214-631-12
; Sequence 12, Application US/09214631
; Patent No. 6413730
; GENERAL INFORMATION:
; APPLICANT: Holland, Sacha
; APPLICANT: Mbamalu, Geraldine
; APPLICANT: Pawson, Tony
; TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED
; TITLE OF INVENTION: BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR
; TITLE OF INVENTION: TYROSINE KINASES
; FILE REFERENCE: 11757.23USWO
; CURRENT APPLICATION NUMBER: US/09/214,631
; CURRENT FILING DATE: 1999-03-12
; EARLIER APPLICATION NUMBER: PCT/CA97/00473
; EARLIER FILING DATE: 1997-07-04
; EARLIER APPLICATION NUMBER: 60/021,272
; EARLIER FILING DATE: 1996-07-05
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 82
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-214-631-12

  Query Match          7.8%;  Score 191.5; DB 4;  Length 82;
  Best Local Similarity 70.6%;  Pred. No. 3.6e-09;
  Matches  36;  Conservative  4; Mismatches  8; Indels  3; Gaps  1;

Qy  288 AEPGELGIALRGGGAADPPFCPHYEKVSGDYGHPVYIVQDGPPQOSPPNIYY 338
            ::| :.| || |  || ||||||||||||||||  |||||
Db   33 SEPSDIIIPLR---TADSVFCPHYEKVSGDYGHPVYIVQEMPPQSPANIYY 80
```

```
RESULT 20
US-08-299-567-5
Sequence 5, Application US/08299567
Patent No. 5747033
GENERAL INFORMATION:
APPLICANT: Davis, et al.
TITLE OF INVENTION: METHOD OF ENHANCING THE BIOLOGICAL
TITLE OF INVENTION: ACTIVITY OF EPH FAMILY LIGANDS
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Regeneron Pharmaceuticals, Inc.
STREET: 777 Old Saw Mill River Road
CITY: Tarrytown
STATE: New York
COUNTRY: U.S.A.
ZIP: 10591-6707
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/299,567
FILING DATE: 01-SEP-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kempler, Gail M.
REGISTRATION NUMBER: 32,143
REFERENCE/DOCKET NUMBER: REG 290
TELECOMMUNICATION INFORMATION:
TELEPHONE: 914-345-7400
TELEFAX: 914-345-7721
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 234 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
```

```
US-08-299-567-5

  Query Match          7.3%;  Score 179;  DB 1;  Length 234;
  Best Local Similarity 28.4%;  Pred. No. 1.6e-07;
  Matches  64;  Conservative  22;  Mismatches  83;  Indels  56;  Gaps 11;

Qy    7 GPGGVRVGALILLGVLLGLVSGLSLEPVTYWNSANKRFQAEGGYVLYPQIGDRLDLLCP--R 64
        ||| |::||:|| :||: :  |:  |  | | | |: |   | :|  |  ||:|||| | |
Db   24 GPG-----------GALG----NRHAVYWNSSNQHLRRE-GYTVQVNVNDYLDIYCPHYN 67

Qy   65 ARPPGPHSSSPNYEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDL---DLRFTIKPQSYS 121
        ||| |  |:| |:  :|   |:| | |  : | | || ||||| |       |||:| :|
Db   68 SSGAGPGPGGAEQYLYMVSRNGYRTCNASQGFK-RWECNRPHAPHSPIKFSRKFQRYS 126

Qy  122 PNLWGHEFRSHHDYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKPV 181
        | | | | ::  |: :|   | |:    | |   | |    | |: |:|   | ||||
Db  127 AFSLGYEFHAGHSYYISTPTHNLH------WKCLR--MKVFVCCASTSHSG---EKPV 174

Qy  182 SEMP----------MERDRGAAHSIE-------PGKENLP 204
        :| |        :|          |              |:||||
Db  175 PTLPQFTMGPNVKINVLEDFEGENPQVPKLEKSISGTSPKREHLP 219
```

```
RESULT 21
US-08-240-124-2
Sequence 2, Application US/08240124
Patent No. 5516658
GENERAL INFORMATION:
APPLICANT: BECKMANN, M. P.
APPLICANT: CERRETTI, DOUGLAS P.
TITLE OF INVENTION: CYTOKINE THAT BINDS THE CELL SURFACE
TITLE OF INVENTION: RECEPTOR HEK
NUMBER OF SEQUENCES: 4
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/240,124
FILING DATE:
CORRESPONDENCE ADDRESS:
ADDRESSEE: IMMUNEX CORPORATION
STREET: 51 UNIVERSITY STREET
CITY: SEATTLE
STATE: WASHINGTON
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Apple System 7.1
SOFTWARE: Microsoft Word for Apple, Version 5.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/240,124
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,132
FILING DATE: 03-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/114,426
FILING DATE: 30-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/109,745
FILING DATE: 20-AUG-1993
ATTORNEY/AGENT INFORMATION:
NAME: SEESE, KATHRYN A.
REGISTRATION NUMBER: 32,172
REFERENCE/DOCKET NUMBER: 2814-C
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 587-0430
TELEFAX: (206) 233-0644
TELEX: 756822
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 238 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-240-124-2
```

```
RESULT 22
US-08-453-943-2
; Sequence 2, Application US/08453943
; Patent No. 5738844
; GENERAL INFORMATION:
;   APPLICANT: BECKMANN, M. P.
;   APPLICANT: CERRETTI, DOUGLAS P.
;   TITLE OF INVENTION: CYTOKINE THAT BINDS THE CELL SURFACE
;   TITLE OF INVENTION: RECEPTOR HEK
;   NUMBER OF SEQUENCES: 4
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: IMMUNEX CORPORATION
;     STREET: 51 UNIVERSITY STREET
;     CITY: SEATTLE
;     STATE: WASHINGTON
;     COUNTRY: USA
;     ZIP: 98101
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: Apple Macintosh
;     OPERATING SYSTEM: Apple System 7.1
;     SOFTWARE: Microsoft Word for Apple, Version 5.1a
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/453,943
;     FILING DATE: 30-MAY-1995
;   PRIOR APPLICATION DATA:
;     CLASSIFICATION: 530
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 08/240,124
;     FILING DATE: 09-MAY-1994
;     APPLICATION NUMBER: US 08/161,132
;     FILING DATE: 03-DEC-1993
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 08/114,426
;     FILING DATE: 30-AUG-1993
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 08/109,745
;     FILING DATE: 20-AUG-1993
;   ATTORNEY/AGENT INFORMATION:
;     NAME: SEESE, KATHRYN A.
;     REGISTRATION NUMBER: 32,172
;     REFERENCE/DOCKET NUMBER: 2814-C
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (206) 587-0430
;     TELEFAX: (206) 233-0644
;     TELEX: 756822
; INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 238 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
;   MOLECULE TYPE: protein

Query Match
  Best Local Similarity   28.4%;   Score 179;   DB 1;   Length 238;
  Matches  65;  Conservative  24;  Mismatches  80;  Indels  60;  Gaps  12;

Qy     7 GPGGVRVGALLLLGVLGLVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDLLCP--R  64
          ||||||   :||:      |||::| :|||::|:|:|:| : ::| :: ::||| : |
Db    24 GPG------GALG------NRHAVYWNSSNQHLRRE-GYTVQVNVNDYLDIYCPHYN     67

Qy    65 ARPPGPHSSP----NYEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDL---DLRFTIKF 117
          : |:   :       ||||| | : ||:|| |:|:| || |  ||: |      |:|:|
Db    68 SSGVGPGAGPGGGGAEQVVLYWVSRNGYRTCNASQGFK-RWECNRPHAPHSPIKFSEKF  126

Qy   118 QEYSPNLWGHEFRSHHDYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVP 177
          | |:  | |:||: | :|||: :| :  :     :  :| |:||| : |: :|:|
Db   127 QRYSAFSLGYEFHAGHEYYYISTPTHNLH------WKCLR--MKVFVCCASTSHSG---  174

Qy   178 RKPVSEMP----------MERDRGAAHSLE-----PGKENLP                  204
          .|||  :           ||||:                :|:|:||
Db   175 EKPVPTLPQFTMGPNVKINVLEDFEGENPQVPKLEKSISGTSPKREHLP           223
```

```
RESULT 23
US-09-057-121-2
; Sequence 2, Application US/09057121
; Patent No. 5969110
; GENERAL INFORMATION:
;   APPLICANT: BECKMANN, M. P.
;   APPLICANT: CERRETTI, DOUGLAS P.
;   TITLE OF INVENTION: CYTOKINE THAT BINDS THE CELL SURFACE
;   TITLE OF INVENTION: RECEPTOR HEK
;   NUMBER OF SEQUENCES: 4
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: IMMUNEX CORPORATION
;     STREET: 51 UNIVERSITY STREET
;     CITY: SEATTLE
;     STATE: WASHINGTON
;     COUNTRY: USA
;     ZIP: 98101
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: Apple Macintosh
;     OPERATING SYSTEM: Apple System 7.1
;     SOFTWARE: Microsoft Word for Apple, Version 5.1a
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/09/057,121
;     FILING DATE:
;   PRIOR APPLICATION DATA:
;     CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: 08/240,124
;     FILING DATE:
;     APPLICATION NUMBER: US 08/161,132
;     FILING DATE: 03-DEC-1993
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 08/114,426
;     FILING DATE: 30-AUG-1993
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 08/109,745
;     FILING DATE: 20-AUG-1993
;   ATTORNEY/AGENT INFORMATION:
;     NAME: SEESE, KATHRYN A.
;     REGISTRATION NUMBER: 32,172
;     REFERENCE/DOCKET NUMBER: 2814-C
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (206) 587-0430
;     TELEFAX: (206) 233-0644
;     TELEX: 756822
; INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 238 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear

Query Match
  Best Local Similarity   28.4%;   Score 179;   DB 1;   Length 238;
  Matches  65;  Conservative  24;  Mismatches  80;  Indels  60;  Gaps  12;

Qy     7 GPGGVRVGALLLLGVLGLVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDLLCP--R  64
          ||||||   :||:      |||::| :|||::|:|:|:| : ::| :: ::||| : |
Db    24 GPG------GALG------NRHAVYWNSSNQHLRRE-GYTVQVNVNDYLDIYCPHYN     67

Qy    65 ARPPGPHSSP----NYEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDL---DLRFTIKF 117
          : |:   :       ||||| | : ||:|| |:|:| || |  ||: |      |:|:|
Db    68 SSGVGPGAGPGGGGAEQVVLYWVSRNGYRTCNASQGFK-RWECNRPHAPHSPIKFSEKF  126

Qy   118 QEYSPNLWGHEFRSHHDYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVP 177
          | |:  | |:||: | :|||: :| :  :     :  :| |:||| : |: :|:|
Db   127 QRYSAFSLGYEFHAGHEYYYISTPTHNLH------WKCLR--MKVFVCCASTSHSG---  174

Qy   178 RKPVSEMP----------MERDRGAAHSLE-----PGKENLP                  204
          .|||  :           ||||:                :|:|:||
Db   175 EKPVPTLPQFTMGPNVKINVLEDFEGENPQVPKLEKSISGTSPKREHLP           223
```

; MOLECULE TYPE: protein
US-09-057-121-2

Query Match
US-09-057-121-2
Best Local Similarity   28.4%;   Score 179; DB 2; Length 238;
Matches 65; Conservative 24; Mismatches 80; Indels 60; Gaps 12;

```
Qy    7 GPGGVRGALLLLGVLGLVSGLSLEPVTWNSANKRFQAEGGYVLLYPQIGDRLDLLCP--R  64
         ||||| ||||||||| ||| ||:||||||||||||||||||||||||||||||||||| |
Db   24 GPG------GALG-----NRHAVYWNSSNQHLRRE-GYTVQVNVNDYLDIYCPHYN  67

Qy   65 ARPPGPHSSP----NYEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDL---DLRFTIKF 117
          || ||||        ||||||||| :|| :|:|| ||||||||| |:||    |||||||
Db   68 SSGVGPGAGPGGGGAEQYVLYMVSRNGYRTCNASQGFK-RWECNRPHAPHSPIKFSEKF 126

Qy  118 QEYSPNLWGHEFRSHHDYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVP 177
          : |||| ||||| ||| ||| ||| |||       |:|| : |||:|| ||
Db  127 QRYSAFSLGYEFHAGHEYYYISTPTHNLH------WKCLR--MKVFVCCASTSHSG--- 174

Qy  178 RKPVSEMP-----------------MERDRGAAHSLE---------PGKENLP 204
          :|:|||||                 ||||||||| |          ||||| |
Db  175 EKPVPTLPQFTMGPNVKINVLEDFEGENPQVPKLEKSISGTSPKREHLP 223
```

RESULT 24
US-09-358-734-2
; Sequence 2, Application US/09358734
; Patent No. 6274117
; GENERAL INFORMATION:
; APPLICANT: BECKMANN, M. P.
; APPLICANT: CERRETTI, DOUGLAS P.
; TITLE OF INVENTION: CYTOKINE THAT BINDS THE CELL SURFACE
; TITLE OF INVENTION: RECEPTOR HEK
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: IMMUNEX CORPORATION
; STREET: 51 UNIVERSITY STREET
; CITY: SEATTLE
; STATE: WASHINGTON
; COUNTRY: USA
; ZIP: 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Apple System 7.1
; SOFTWARE: Microsoft Word for Apple, Version 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/358,734
; FILING DATE:
; PRIOR APPLICATION DATA:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/240,124
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/114,426
; FILING DATE: 30-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/109,745
; FILING DATE: 20-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: SEESE, KATHRYN A.
; REGISTRATION NUMBER: 32,172
; REFERENCE/DOCKET NUMBER: 2814-C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 587-0430
; TELEFAX: (206) 233-0644
; TELEX: 756822
INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 238 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

US-09-358-734-2

Query Match
US-09-358-734-2
Best Local Similarity   28.4%;   Score 179; DB 4; Length 238;
Matches 65; Conservative 24; Mismatches 80; Indels 60; Gaps 12;

```
Qy    7 GPGGVRGALLLLGVLGLVSGLSLEPVTWNSANKRFQAEGGYVLLYPQIGDRLDLLCP--R  64
         ||||| ||||||||| ||| ||:||||||||||||||||||||||||||||||||||| |
Db   24 GPG------GALG-----NRHAVYWNSSNQHLRRE-GYTVQVNVNDYLDIYCPHYN  67

Qy   65 ARPPGPHSSP----NYEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDL---DLRFTIKF 117
          || ||||        ||||||||| :|| :|:|| ||||||||| |:||    |||||||
Db   68 SSGVGPGAGPGGGGAEQYVLYMVSRNGYRTCNASQGFK-RWECNRPHAPHSPIKFSEKF 126

Qy  118 QEYSPNLWGHEFRSHHDYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVP 177
          : |||| ||||| ||| ||| ||| |||       |:|| : |||:|| ||
Db  127 QRYSAFSLGYEFHAGHEYYYISTPTHNLH------WKCLR--MKVFVCCASTSHSG--- 174

Qy  178 RKPVSEMP-----------------MERDRGAAHSLE---------PGKENLP 204
          :|:|||||                 ||||||||| |          ||||| |
Db  175 EKPVPTLPQFTMGPNVKINVLEDFEGENPQVPKLEKSISGTSPKREHLP 223
```

RESULT 25
US-08-299-567-7
; Sequence 7, Application US/08299567
; Patent No. 5547033
; GENERAL INFORMATION:
; APPLICANT: Davis, et al.
; TITLE OF INVENTION: METHOD OF ENHANCING THE BIOLOGICAL
; TITLE OF INVENTION: ACTIVITY OF EPH FAMILY LIGANDS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Regeneron Pharmaceuticals, Inc.
; STREET: 777 Old Saw Mill River Road
; CITY: Tarrytown
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10591-6707
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/299,567
; FILING DATE: 01-SEP-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kempler, Gail M.
; REGISTRATION NUMBER: 32,143
; REFERENCE/DOCKET NUMBER: REG 290
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 914-345-7400
; TELEFAX: 914-345-7721
INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 135 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein

US-08-299-567-7

Query Match
Best Local Similarity   31.9%;   Score 176.5; DB 1; Length 135;
Matches 43; Conservative 14; Mismatches 73; Indels 5; Gaps 2;

```
Qy   33 VYWNSANKRFQAEGGYVLLYPQIGDRLDLLCP--RARPPGPHSSPSNYEFYKLYLVGGAQGR  90
         | || |  || | ||| ||| | |||||| ||                  |||||| |
Db    1 VXWNSSNPKFXRXEGYTIVXXNDYLDIICPHYEXXXXXXXAGXXXECYXLYLYVXXEXXXX  60

Qy   91 RCEAPPAPNLLLTCDRP---DLDLRFTIKFQEYSPNLWGHEFRSHHDYYIIATSDGTREG 147
         |:|| || |||| |||     |||||||||
Db   60 ... 147
```

Db   61 XCXXXSXXXVRWXCNRPXAVHXPIXFSEKFQRFSPFXLGXEFKXGHXYYYISTXXXPXHX 120

Qy 148 LESLQGGVCLTRGMK 162

Db 121 LEXRXXXXCXLRXMK 135

RESULT 26
US-09-609-324A-2
Sequence 2, Application US/09609324A
Patent No. RE37582
GENERAL INFORMATION:
APPLICANT: CERRETTI, Douglas P.
TITLE OF INVENTION: CYTOKINE DESIGNATED LERK-6
FILE REFERENCE: A7772
CURRENT APPLICATION NUMBER: US/09/609,324A
CURRENT FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER:
PRIOR FILING DATE: 1997-08-29
PRIOR APPLICATION NUMBER: 08/538,709
PRIOR FILING DATE: 1995-10-03
PRIOR APPLICATION NUMBER: 08/920,440
PRIOR FILING DATE:
PRIOR APPLICATION NUMBER: 08/318,393
PRIOR FILING DATE: 1994-10-05
NUMBER OF SEQ ID NOS: 10
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 184
TYPE: PRT
ORGANISM: LERK-6
US-09-609-324A-2

Query Match   7.2%; Score 176; DB 1; Length 184;
Best Local Similarity 29.3%; Pred. No. 2e-07;
Matches 58; Conservative 19; Mismatches 69; Indels 52; Gaps 7;

Qy   33 VYWNSANKRFQAE------GGYVLYPQIGDRLDLLCPRARPPGPHSSPNYEFYKLYLVGGA 87
         ||| :| || : |||:|: ||:|||: |||:|:||:
Db   10 VYWNRSNPRFQVSAVGDGGGYTVEVSINDYLDIYCPHYGAPLP-PAERMERYILYMVNGE 68

Qy   88 QGRRCEAPPAPNLLLTCDRPDL---DLRFTIKFQEYSPNLWGHEFRSHHDYYIIATSDGT 144
         |:| |:|||:|:||:||| ||||:|:||:||||
Db   69 GHASCDHRQRGFKRWECNRPAAPGGPLKFSEKFQLFTPFSLGFEFRPGHEYYYISATP-- 126

Qy 145 REGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKPVSEMPMERDRGAAHSLEPGKENLP 204
         |:| |||:|| :||:|: |||| |: | : | : ||:|
Db 127 ----PNLVDRPCLR--LKYVR--------------------------------PTNETLY 149

Qy 205 GDP----TSNATSRGAEG 218
            |||: :
Db 150 EAPEPIFTSNSSCSGLGG 167

RESULT 27
US-08-920-440B-2
Sequence 2, Application US/0892040440B
Patent No. 5913905
GENERAL INFORMATION:
APPLICANT: Cerretti, Douglas P.
TITLE OF INVENTION: Cytokine Designated LERK-6
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Immunex Corporation
STREET: 51 University Street
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Power Macintosh
OPERATING SYSTEM: System 7.6
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/920,440B
FILING DATE: 29-AUG-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Henry, Janis C.
REGISTRATION NUMBER: 34,347
REFERENCE/DOCKET NUMBER: 2826-B
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 587-0430
TELEFAX: (206) 233-0644
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 184 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-920-440B-2

Query Match   7.2%; Score 176; DB 2; Length 184;
Best Local Similarity 29.3%; Pred. No. 2e-07;
Matches 58; Conservative 19; Mismatches 69; Indels 52; Gaps 7;

Qy   33 VYWNSANKRFQAE------GGYVLYPQIGDRLDLLCPRARPPGPHSSPNYEFYKLYLVGGA 87
         ||| :| || : |||:|: ||:|||: |||:|:||:
Db   10 VYWNRSNPRFQVSAVGDGGGYTVEVSINDYLDIYCPHYGAPLP-PAERMERYILYMVNGE 68

Qy   88 QGRRCEAPPAPNLLLTCDRPDL---DLRFTIKFQEYSPNLWGHEFRSHHDYYIIATSDGT 144
         |:| |:|||:|:||:||| ||||:|:||:||||
Db   69 GHASCDHRQRGFKRWECNRPAAPGGPLKFSEKFQLFTPFSLGFEFRPGHEYYYISATP-- 126

Qy 145 REGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKPVSEMPMERDRGAAHSLEPGKENLP 204
         |:| |||:|| :||:|: |||| |: | : | : ||:|
Db 127 ----PNLVDRPCLR--LKYVR--------------------------------PTNETLY 149

Qy 205 GDP----TSNATSRGAEG 218
            |||: :
Db 150 EAPEPIFTSNSSCSGLGG 167

RESULT 28
US-09-173-492-2
Sequence 2, Application US/09173492
Patent No. 6194172
GENERAL INFORMATION:
APPLICANT: Cerretti, Douglas P.
TITLE OF INVENTION: Cytokine Designated LERK-6
NUMBER OF SEQUENCES: 10
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/173,492
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/920,440
FILING DATE: 29-AUG-1997
ATTORNEY/AGENT INFORMATION:
NAME: Henry, Janis C.
REGISTRATION NUMBER: 34,347
REFERENCE/DOCKET NUMBER: 2826-B
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 587-0430

```
; TELEFAX: (206) 233-0644
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 184 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-173-492-2

Query Match          7.2%;  Score 176;  DB 4;  Length 184;
Best Local Similarity   29.3%;  Pred. No. 2e-07;
Matches   58;  Conservative  19;  Mismatches  69;  Indels  52;  Gaps   7;

Qy   33 VYWNSANKFQAE-----GGYVLYPQIGDRLDLLCPRARPPGPHSSPNYEFYKLYLVGGA 87
           ||||  |||:|       ||:|||||:||||||||:|||||:| |  : |||:|||:
Db   10 VYWNRSNPRFQVSAVGDGGGYTVEVSINDYLDIYCPHYGAPLP-PAERMERYILYMVNGE 68

Qy   88 QGRRCEAPPAPNLLLTCDRPDL---DLRFTIKFQEYSPNLWGHEFRSHHDYYIIATSDGT 144
            |:  |||||:|||||||: |   |||:||||: ||::|||||:||||:|||||||
Db   69 GHASCDHRQGFKRWECNRPAAPGGPLKFSEKFQLFTPFSLGFEFRPGHEYYYISATP-- 126

Qy  145 REGLESLQGGVCLTRGMKVLLRVGQSPRGAVPRKPVSEMPMERDRGAAHSLBPGKENLP 204
            |   | ||||: | ||  |  |                             :|| |
Db  127 ----PNLVDRPCLR--LKYVVR-------------------------------PINETLY 149

Qy  205 GDP-----TSNATSRGAEG 218
             |:      |||:  |||
Db  150 EAPEPIFTSNSSCSGLGG 167

RESULT 29
US-09-173-133-2
; Sequence 2, Application US/09173133
; Patent No. 6232447
; GENERAL INFORMATION:
; APPLICANT: Cerretti, Douglas P.
; TITLE OF INVENTION: Cytokine Designated LERK-6
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Immunex Corporation
; STREET: 51 University Street
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Power Macintosh
; OPERATING SYSTEM: System 7.6
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/173,133
; FILING DATE:
; PRIOR APPLICATION DATA:
; CLASSIFICATION:
; APPLICATION NUMBER: US/08/920,440
; FILING DATE: 29-AUG-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Henry, Janis C.
; REGISTRATION NUMBER: 34,347
; REFERENCE/DOCKET NUMBER: 2826-B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 587-0430
; TELEFAX: (206) 233-0644
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 184 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-173-133-2

Query Match          7.2%;  Score 176;  DB 4;  Length 184;
```

```
Best Local Similarity   29.3%;  Pred. No. 2e-07;
Matches   58;  Conservative  19;  Mismatches  69;  Indels  52;  Gaps   7;

Qy   33 VYWNSANKFQAE-----GGYVLYPQIGDRLDLLCPRARPPGPHSSPNYEFYKLYLVGGA 87
           ||||  |||:|       ||:|||||:||||||||:|||||:| |  : |||:|||:
Db   10 VYWNRSNPRFQVSAVGDGGGYTVEVSINDYLDIYCPHYGAPLP-PAERMERYILYMVNGE 68

Qy   88 QGRRCEAPPAPNLLLTCDRPDL---DLRFTIKFQEYSPNLWGHEFRSHHDYYIIATSDGT 144
            |:  |||||:|||||||: |   |||:||||: ||::|||||:||||:|||||||
Db   69 GHASCDHRQGFKRWECNRPAAPGGPLKFSEKFQLFTPFSLGFEFRPGHEYYYISATP-- 126

Qy  145 REGLESLQGGVCLTRGMKVLLRVGQSPRGAVPRKPVSEMPMERDRGAAHSLBPGKENLP 204
            |   | ||||: | ||  |  |                             :|| |
Db  127 ----PNLVDRPCLR--LKYVVR-------------------------------PINETLY 149

Qy  205 GDP-----TSNATSRGAEG 218
             |:      |||:  |||
Db  150 EAPEPIFTSNSSCSGLGG 167

RESULT 30
US-09-165-533-2
; Sequence 2, Application US/09165533
; Patent No. 6268482
; GENERAL INFORMATION:
; APPLICANT: Cerretti, Douglas P.
; TITLE OF INVENTION: Cytokine Designated as LERK-6
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Immunex Corporation
; STREET: 51 University Street
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: System 7.1
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/165,533
; FILING DATE:
; PRIOR APPLICATION DATA:
; CLASSIFICATION:
; APPLICATION NUMBER: 08/538,709
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Malaska, Stephen L.
; REGISTRATION NUMBER: 32,655
; REFERENCE/DOCKET NUMBER: 2826
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 587-0430
; TELEFAX: (206) 233-0644
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 184 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-165-533-2

Query Match          7.2%;  Score 176;  DB 4;  Length 184;
Best Local Similarity   29.3%;  Pred. No. 2e-07;
Matches   58;  Conservative  19;  Mismatches  69;  Indels  52;  Gaps   7;

Qy   33 VYWNSANKFQAE-----GGYVLYPQIGDRLDLLCPRARPPGPHSSPNYEFYKLYLVGGA 87
           ||||  |||:|       ||:|||||:||||||||:|||||:| |  : |||:|||:
Db   10 VYWNRSNPRFQVSAVGDGGGYTVEVSINDYLDIYCPHYGAPLP-PAERMERYILYMVNGE 68

Qy   88 QGRRCEAPPAPNLLLTCDRPDL---DLRFTIKFQEYSPNLWGHEFRSHHDYYIIATSDGT 144
            |:  |||||:|||||||: |   |||:||||: ||::|||||:||||:|||||||
Db   69 GHASCDHRQGFKRWECNRPAAPGGPLKFSEKFQLFTPFSLGFEFRPGHEYYYISATP-- 126
```

RESULT 31
PCT-US95-12779-2
; Sequence 2, Application PCT/US9512779
; GENERAL INFORMATION:
; APPLICANT: Cerretti, Douglas P.
; TITLE OF INVENTION: Cytokine Designated LERK-6
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Immunex Corporation
; STREET: 51 University Street
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: System 7.1
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/12779
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Malaska, Stephen L.
; REGISTRATION NUMBER: 32,655
; REFERENCE/DOCKET NUMBER: 2826-WO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 587-0430
; TELEFAX: (206) 233-0644
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 184 amino acid
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US95-12779-2

Query Match   7.2%; Score 176; DB 5; Length 184;
Best Local Similarity   29.3%; Pred. No. 2e-07;
Matches 58; Conservative 19; Mismatches 69; Indels 52; Gaps 7;

Qy 33 VYWNSANKRFQAE-----GGYVLYPQIGDRLDLLCPRARPPGPHSSPNYEFYKLYLVGGA 87
Db 10 VYWNRSNPRFQVSAVGDGGGYTVEVSINDYLDIYCPHYGAPLP-PAERMERYILYMVNGE 68

Qy 88 QGRRCEAPPAPNLLLTCDRPDL---DLRFTIKFQEYSPNLWGHEFRSHHDYYIIATSDGT 144
Db 69 GHASCDHRQRGFKRWECNRPAAPGGPLKFSEKFQLFTPFSLGFBFRPGHEYYISATP-- 126

Qy 145 REGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKPVSEMPMERDRGAAHSLEPGKENLP 204
Db 127 ----PNLVDRPCLR--LKYVVR------------------------PTNETLY 149

Qy 205 GDP----TSNATSRGAEG 218
Db 150 EAPEPIFTSNSSCSGLGG 167

RESULT 32
PCT-US95-15781-2
; Sequence 2, Application PCT/US9515781
; GENERAL INFORMATION:

RESULT 33
US-08-455-001-2
; Sequence 2, Application US/08455001
; Patent No. 5795734
; GENERAL INFORMATION:
; APPLICANT: Flanagan, John G.
; APPLICANT: Cheng, Hwai-Jong
; TITLE OF INVENTION: EPH Receptor Ligands, and Uses Related
; TITLE OF INVENTION: Thereto
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:

; APPLICANT: Cerretti, Douglas P.
; TITLE OF INVENTION: Cytokine Designated Lerk-7
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Immunex Corporation
; STREET: 51 University Street
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: System 7.1
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/15781
; FILING DATE: 05-DEC-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/351,025
; FILING DATE: 06-DEC-1994
; CLASSIFICATION:
; APPLICATION NUMBER: US 08/396,946
; FILING DATE: 01-MAR-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Anderson, Kathryn A.
; REGISTRATION NUMBER: 32,172
; REFERENCE/DOCKET NUMBER: 2829-WO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 587-0430
; TELEFAX: (206) 233-0644
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 184 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US95-15781-2

Query Match   7.2%; Score 176; DB 5; Length 184;
Best Local Similarity   29.3%; Pred. No. 2e-07;
Matches 58; Conservative 19; Mismatches 69; Indels 52; Gaps 7;

Qy 33 VYWNSANKRFQAE-----GGYVLYPQIGDRLDLLCPRARPPGPHSSPNYEFYKLYLVGGA 87
Db 10 VYWNRSNPRFQVSAVGDGGGYTVEVSINDYLDIYCPHYGAPLP-PAERMERYILYMVNGE 68

Qy 88 QGRRCEAPPAPNLLLTCDRPDL---DLRFTIKFQEYSPNLWGHEFRSHHDYYIIATSDGT 144
Db 69 GHASCDHRQRGFKRWECNRPAAPGGPLKFSEKFQLFTPFSLGFBFRPGHEYYISATP-- 126

Qy 145 REGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKPVSEMPMERDRGAAHSLEPGKENLP 204
Db 127 ----PNLVDRPCLR--LKYVVR------------------------PTNETLY 149

Qy 205 GDP----TSNATSRGAEG 218
Db 150 EAPEPIFTSNSSCSGLGG 167

```
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street
STATE: MA
CITY: Boston
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII (text)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/455,001
FILING DATE: 31 MAY 1995
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Vincent, Matthew P.
REGISTRATION NUMBER: 36,709
REFERENCE/DOCKET NUMBER: HMI-011CP2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 209 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-455-001-2

Query Match      7.2%; Score 176; DB 1; Length 209;
Best Local Similarity  29.3%; Pred. No. 2.4e-07;
Matches  58; Conservative  19; Mismatches  69; Indels  52; Gaps  7;

Qy   33 VYWNSANKRFQAE------GGYVLYPQIGDRLDLLCPRARPPGPHSSPNYEFYKLYLVGGA  87
        ||||  :| |||  |  : |||||: ||:
Db   35 VTWNRSNPRFQVSAVGDGGGYTVEVSINDYLDIYCPHYGAPLP-PAERMERYILYMVNGE  93

Qy   88 QGRRCEAPPAPNLLLTCDRPDL---DLRFTIKFQEYSPNLWGHEFRSHHDYYIIATSDGT 144
        ||  |  : :|  :  ||  |  | :: |  : :
Db   94 GHASCDHRQRGFKRWECNRPAAPGGPLKFSEKFQLFTFSLGFEFRPGHEYYISATP-- 151

Qy  145 REGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKPVSEMPERDRGAAHSLEPGKENLP 204
        | :|   |   |  :| :  :|
Db  152 ----PNLVDRPCLR--LKYVVR-------------------------PTNETLY 174

Qy  205 GDP----TSNATSRGAEG 218
        ||:   ||:::
Db  175 EAPEPIFTSNSSCSGLGG 192

RESULT 34
US-08-308-814-2
Sequence 2, Application US/08308814
Patent No. 6268476
GENERAL INFORMATION:
APPLICANT: Flanagan, John G.
APPLICANT: Cheng, Hwai-Jong
TITLE OF INVENTION: EPH Receptor Ligands, and Uses Related
  Thereto
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII (text)
```

```
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/308,814
FILING DATE: 19-SEP-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Vincent, Matthew P.
REGISTRATION NUMBER: 36,709
REFERENCE/DOCKET NUMBER: HMI-011
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 209 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-308-814-2

Query Match      7.2%; Score 176; DB 4; Length 209;
Best Local Similarity  29.3%; Pred. No. 2.4e-07;
Matches  58; Conservative 19; Mismatches  69; Indels  52; Gaps  7;

Qy   33 VYWNSANKRFQAE------GGYVLYPQIGDRLDLLCPRARPPGPHSSPNYEFYKLYLVGGA  87
        ||||  :| |||  |  : |||||: ||:
Db   35 VTWNRSNPRFQVSAVGDGGGYTVEVSINDYLDIYCPHYGAPLP-PAERMERYILYMVNGE  93

Qy   88 QGRRCEAPPAPNLLLTCDRPDL---DLRFTIKFQEYSPNLWGHEFRSHHDYYIIATSDGT 144
Db   94 GHASCDHRQRGFKRWECNRPAAPGGPLKFSEKFQLFTFSLGFEFRPGHEYYISATP-- 151

Qy  145 REGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKPVSEMPERDRGAAHSLEPGKENLP 204
Db  152 ----PNLVDRPCLR--LKYVVR-------------------------PTNETLY 174

Qy  205 GDP----TSNATSRGAEG 218
Db  175 EAPEPIFTSNSSCSGLGG 192

RESULT 35
PCT-US95-11869-2
Sequence 2, Application PCT/US9511869
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: EPH Receptor Ligands, and Uses Related
  Thereto
NUMBER OF SEQUENCES: 5
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE:
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/11869
FILING DATE: 19-SEP-1995
ATTORNEY/AGENT INFORMATION:
NAME: Vincent, Matthew P.
REGISTRATION NUMBER: 36,709
REFERENCE/DOCKET NUMBER: HMI-011CPPC
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 209 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-11869-2

Query Match      7.2%; Score 176; DB 5; Length 209;
Best Local Similarity  29.3%; Pred. No. 2.4e-07;
```

```
Matches    58; Conservative    19; Mismatches    69; Indels    52; Gaps    7;

Qy   33 VYWNSANKRFQAE------GGYVLYPQIGDRLDLLCPRARPPGPHSSPNYEFYKLYLVGGA 87
           |||:.|::|            :.||||:||||:||||:||||||||:||||||||||
Db   35 VYWNRSNPRFQVSAVGDGGGYTVEVSINDYLDIYCPHYGAPLP-PAERMERYILYMVNGE 93

Qy   88 QGRRCEAPPAPNLLLTCDRPDL---DLRFTIKFQEYSPNLWGHEFRSHHDYIIATSDGT 144
          :||  |||:|  :||  |:||    |:|.:||:|||||||||||:||||||||:
Db   94 GHASCDHRQRGFKRWECNRPAAPGGPLKFSEKFQLFTPFSLGFEFRPGHEYYISATP-- 151

Qy  145 REGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKPVSEMPERDRGAAHSLBPGKENLP 204
Db  152 ---PNLVDRPCLR--LKVYVR-----------------------------PTNETLY 174

Qy  205 GDP-----TSNATSRGAEG 218
              |||:||
Db  175 EAPEPIPTSNSSCSGLGG 192
```

RESULT 36
US-09-609-324A-10
; Sequence 10, Application US/0960324A
; Patent No. RE37582
; GENERAL INFORMATION:
; APPLICANT: CERRETTI, Douglas P.
; TITLE OF INVENTION: CYTOKINE DESIGNATED LERK-6
; FILE REFERENCE: A7772
; CURRENT APPLICATION NUMBER: US/09/609,324A
; CURRENT FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: 08/920,440
; PRIOR FILING DATE: 1997-08-29
; PRIOR APPLICATION NUMBER: 08/538,709
; PRIOR FILING DATE: 1995-10-03
; PRIOR APPLICATION NUMBER: 08/318,393
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 10
; LENGTH: 213
; TYPE: PRT
; ORGANISM: LERK-6
US-09-609-324A-10

```
Query Match       7.2%; Score 175.5; DB 1; Length 213;
Best Local Similarity   36.8%; Pred. No. 2.7e-07;
Matches   43; Conservative   14; Mismatches   51; Indels    9; Gaps    3;

Qy   33 VYWNSANKRFQA------EGGYVLYPQIGDRLDLLCPRARPPGPHSSPNYEFYKLYLVGGA 87
Db   39 VYWNRSNPRFHAGAGDDGGGYTVEVSINDYLDIYCPHYGAPLP-PAERMEHYVLYMVNGE 97

Qy   88 QGRRCEAPPAPNLLLTCDRPDL---DLRFTIKFQEYSPNLWGHEFRSHHDYYIIATS 141
Db   98 GHASCDHRQRGFKRWECNRPAAPGGPLKFSEKFQLFTPFSLGFEFRPGHEYYISAT 154
```

RESULT 37
US-08-920-440B-10
; Sequence 10, Application US/08920440B
; Patent No. 5919905
; GENERAL INFORMATION:
; APPLICANT: Cerretti, Douglas P.
; TITLE OF INVENTION: Cytokine Designated LERK-6
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Immunex Corporation
; STREET: 51 University Street
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

; COMPUTER: Apple Power Macintosh
; OPERATING SYSTEM: System 7.6
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/920,440B
; FILING DATE: 29-AUG-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Henry, Janis C.
; REGISTRATION NUMBER: 34,347
; REFERENCE/DOCKET NUMBER: 2826-B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 587-0430
; TELEFAX: (206) 233-0644
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 213 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-920-440B-10

```
Query Match       7.2%; Score 175.5; DB 2; Length 213;
Best Local Similarity   36.8%; Pred. No. 2.7e-07;
Matches   43; Conservative   14; Mismatches   51; Indels    9; Gaps    3;

Qy   33 VYWNSANKRFQA------EGGYVLYPQIGDRLDLLCPRARPPGPHSSPNYEFYKLYLVGGA 87
Db   39 VYWNRSNPRFHAGAGDDGGGYTVEVSINDYLDIYCPHYGAPLP-PAERMEHYVLYMVNGE 97

Qy   88 QGRRCEAPPAPNLLLTCDRPDL---DLRFTIKFQEYSPNLWGHEFRSHHDYYIIATS 141
Db   98 GHASCDHRQRGFKRWECNRPAAPGGPLKFSEKFQLFTPFSLGFEFRPGHEYYISAT 154
```

RESULT 38
US-09-173-492-10
; Sequence 10, Application US/09173492
; Patent No. 6194172
; GENERAL INFORMATION:
; APPLICANT: Cerretti, Douglas P.
; TITLE OF INVENTION: Cytokine Designated LERK-6
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Immunex Corporation
; STREET: 51 University Street
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Power Macintosh
; OPERATING SYSTEM: System 7.6
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/173,492
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/920,440
; FILING DATE: 29-AUG-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Henry, Janis C.
; REGISTRATION NUMBER: 34,347
; REFERENCE/DOCKET NUMBER: 2826-B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 587-0430
; TELEFAX: (206) 233-0644
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 213 amino acids
; TYPE: amino acid

```
;  TOPOLOGY: linear
;  MOLECULE TYPE: protein
US-09-173-492-10

Qy  33  VYWNSANKRFQA-----EGGYVLYPQIGDRLDLLCPRARPPGPHSSPNYEFYKLYLVGGA  87
        ||||:| |||||    |   ||||:||||||||||:|: |  |:|| :|:|: || |:
Db  39  VYWNRSNPRFHAGAGDDGGGYTVEVSINDYLDIYCPHYGAPLP-PAERMEHVLYMVNGE  97

Qy  88  QGRRCEAPPAPNLLLTCDRPDL----DLRFTIKFQEYSPNLWGHEFRSHHDYYIIATS  141
        |:| |  ||  |:||:||| |     |||||||||::|  :|  :| :
Db  98  GHASCDHRQRGFKRWECNRPAAPGGPLKFSEKFQLFTPFSLGFEFRPGHEYYYISAT  154
```

RESULT 39
```
US-09-173-133-10
;  Sequence 10, Application US/0917313
;  Patent No. 6232447
;  GENERAL INFORMATION:
;  APPLICANT: Cerretti, Douglas P.
;  TITLE OF INVENTION: Cytokine Designated LERK-6
;  NUMBER OF SEQUENCES: 10
;  CORRESPONDENCE ADDRESS:
;  ADDRESSEE: Immunex Corporation
;  STREET: 51 University Street
;  CITY: Seattle
;  STATE: WA
;  COUNTRY: USA
;  ZIP: 98101
;  COMPUTER READABLE FORM:
;  MEDIUM TYPE: Floppy disk
;  COMPUTER: Apple Power Macintosh
;  OPERATING SYSTEM: System 7.6
;  SOFTWARE: PatentIn Release #1.0, Version #1.25
;  CURRENT APPLICATION DATA:
;  APPLICATION NUMBER: US/09/173,133
;  FILING DATE:
;  CLASSIFICATION:
;  PRIOR APPLICATION DATA:
;  APPLICATION NUMBER: US/08/920,440
;  FILING DATE: 29-AUG-1997
;  ATTORNEY/AGENT INFORMATION:
;  NAME: Henry, Janis C.
;  REGISTRATION NUMBER: 34,347
;  REFERENCE/DOCKET NUMBER: 2826-B
;  TELECOMMUNICATION INFORMATION:
;  TELEPHONE: (206) 587-0430
;  TELEFAX: (206) 233-0644
;  INFORMATION FOR SEQ ID NO: 10:
;  SEQUENCE CHARACTERISTICS:
;  LENGTH: 213 amino acids
;  TYPE: amino acid
;  TOPOLOGY: linear
;  MOLECULE TYPE: protein
US-09-173-133-10
```

Query Match   36.8%; Score 175.5; DB 4; Length 213;
Best Local Similarity   36.8%; Pred. No. 2.7e-07;
Matches   43; Conservative   14; Mismatches   51; Indels   9; Gaps   3;

```
Qy  33  VYWNSANKRFQA-----EGGYVLYPQIGDRLDLLCPRARPPGPHSSPNYEFYKLYLVGGA  87
Db  39  VYWNRSNPRFHAGAGDDGGGYTVEVSINDYLDIYCPHYGAPLP-PAERMEHVLYMVNGE  97
Qy  88  QGRRCEAPPAPNLLLTCDRPDL----DLRFTIKFQEYSPNLWGHEFRSHHDYYIIATS  141
Db  98  GHASCDHRQRGFKRWECNRPAAPGGPLKFSEKFQLFTPFSLGFEFRPGHEYYYISAT  154
```

RESULT 40

```
US-09-214-631-7
;  Sequence 7, Application US/09214631
;  Patent No. 6413730
;  GENERAL INFORMATION:
;  APPLICANT: Holland, Sacha
;  APPLICANT: Mbamalu, Geraldine
;  APPLICANT: Pawson, Tony
;  TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED
;  TITLE OF INVENTION: BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR
;  TITLE OF INVENTION: TYROSINE KINASES
;  FILE REFERENCE: 11757.23USWO
;  CURRENT APPLICATION NUMBER: US/09/214,631
;  CURRENT FILING DATE: 1999-03-12
;  EARLIER APPLICATION NUMBER: PCT/CA97/00473
;  EARLIER FILING DATE: 1997-07-04
;  EARLIER APPLICATION NUMBER: 60/021,272
;  EARLIER FILING DATE: 1996-07-05
;  NUMBER OF SEQ ID NOS: 13
SEQ ID NO 7
;  LENGTH: 233
;  TYPE: PRT
;  ORGANISM: Homo sapiens
US-09-214-631-7
```

Query Match   27.4%; Score 174.5; DB 4; Length 233;
Best Local Similarity   27.4%; Pred. No. 3.7e-07;
Matches   60; Conservative   24; Mismatches   74; Indels   61; Gaps   11;

```
Qy   20  GVLGLVSGLSLEPVTWNSANKRFQABGGYVLYPQIGDRLDLLCPR----ARPPGPHSSPN   75
Db   27  GALG-----NRHAVYWNSSNQHLRRE-GYTVQVNVNDYLDIYCPHYNSSGAGPGPGGAE   80
Qy   76  YEFYKLYLVG-----GAQGRR---CEAPPAPNLLLTCDRPDLDLRFTIKFQEYSPNLWGH  127
Db   81  YVLYMVSRNGYRTCNASQGFKRWECNRPAPH-------SPIKFSEKFQRYSAFSLGY  131
Qy  128  EFRSHHDYYIIATSDGTRGLESLQGGVCLTRGMKVLRVGQSPRGGAVPRKPVSEMP--  185
Db  132  EFHAGHEYYYISTPTHNLH------WKCLR--MKVFVCCASTSHSG---EKVPTLPQF  179
Qy  186  ------------MERDRGAAHSLE----------PGKENLP  204
Db  180  TMGPNVKINVLEDFEGENPQVPKLEKSISGTSPKREHLP  218
```

RESULT 41
```
US-08-455-001-4
;  Sequence 4, Application US/08455001
;  Patent No. 5795734
;  GENERAL INFORMATION:
;  APPLICANT: Flanagan, John G.
;  APPLICANT: Cheng, Hwai-Jong
;  TITLE OF INVENTION: EPH Receptor Ligands, and Uses Related
;  TITLE OF INVENTION: Thereto
;  NUMBER OF SEQUENCES: 5
;  CORRESPONDENCE ADDRESS:
;  ADDRESSEE: LAHIVE & COCKFIELD
;  STREET: 60 State Street
;  CITY: Boston
;  STATE: MA
;  COUNTRY: USA
;  ZIP: 02109
;  COMPUTER READABLE FORM:
;  MEDIUM TYPE: Floppy disk
;  COMPUTER: IBM PC compatible
;  OPERATING SYSTEM: PC-DOS/MS-DOS
;  SOFTWARE: ASCII (text)
;  CURRENT APPLICATION DATA:
;  APPLICATION NUMBER: US/08/455,001
;  FILING DATE: 31 MAY 1995
;  CLASSIFICATION: 800
;  ATTORNEY/AGENT INFORMATION:
```

```
;  NAME: Vincent, Matthew P.
;  REGISTRATION NUMBER: 36,709
;  REFERENCE/DOCKET NUMBER: HMI-011CIP2
;  TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (617) 227-7400
;   TELEFAX: (617) 227-5941
;
; INFORMATION FOR SEQ ID NO: 4:
;  SEQUENCE CHARACTERISTICS:
;   LENGTH: 200 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-455-001-4

Query Match       7.0%; Score 172; DB 1; Length 200;
Best Local Similarity   35.3%; Pred. No. 4.9e-07;
Matches  49; Conservative  16; Mismatches  58; Indels  16; Gaps  5;

Qy   15 ALILLGVLGLVSGLSLEP-------------VYWNSANKRFQAEGGYVLYPQIGDRLDLLCPRA 65
            ||| ||:|:|                   ||||| |:|  ||
Db    7 AALLAAIVG-VCVWSDDPGKVISDRYAVYWNRSNPRFH-RGDYTVEVSINDYLDIYCPHY 64

Qy   66 RPPGPHSSPNYEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDL---DLRFTIKFQEYSP 122
            ||:|  | |||  |:||  ||:  |  |    ||||| |:  || :|||:||| || |
Db   65 EEPLP--AERMERYLIVMVNYEGHASCDHRQGFKRWECNRPDSPSGGPLKFSEKFQLFTP 122

Qy  123 NLWGHEFRSHHDYYIIATS 141
            |:|| ||||:||||:|:
Db  123 FSLGFEFRPGHEYYISAS 141
```

```
RESULT 42
; Sequence 4, Application PCT/US9511869
; GENERAL INFORMATION:
;  APPLICANT:
;  TITLE OF INVENTION: EPH Receptor Ligands, and Uses Related
;   Thereto
;  NUMBER OF SEQUENCES: 5
;  COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE:
;  CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: PCT/US95/11869
;   FILING DATE: 19-SEP-1995
;  ATTORNEY/AGENT INFORMATION:
;   NAME: Vincent, Matthew P.
;   REGISTRATION NUMBER: 36,709
;   REFERENCE/DOCKET NUMBER: HMI-011CPPC
;  TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (617) 227-7400
;   TELEFAX: (617) 227-5941
;
; INFORMATION FOR SEQ ID NO: 4:
;  SEQUENCE CHARACTERISTICS:
;   LENGTH: 200 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
PCT-US95-11869-4

Query Match       7.0%; Score 172; DB 5; Length 200;
Best Local Similarity   35.3%; Pred. No. 4.9e-07;
Matches  49; Conservative  16; Mismatches  58; Indels  16; Gaps  5;

Qy   15 ALILLGVLGLVSGLSLEP-------------VYWNSANKRFQAEGGYVLYPQIGDRLDLLCPRA 65
            ||| ||:|:|                   ||||| |:|  ||
Db    7 AALLAAIVG-VCVWSDDPGKVISDRYAVYWNRSNPRFH-RGDYTVEVSINDYLDIYCPHY 64

Qy   66 RPPGPHSSPNYEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDL---DLRFTIKFQEYSP 122
            ||:|  | |||  |:||  ||:  |  |    ||||| |:  || :|||:||| || |
Db   65 EEPLP--AERMERYLIVMVNYEGHASCDHRQGFKRWECNRPDSPSGGPLKFSEKFQLFTP 122

Qy  123 NLWGHEFRSHHDYYIIATS 141
            |:|| ||||:||||:|:
Db  123 FSLGFEFRPGHEYYISAS 141
```

```
Qy  123 NLWGHEFRSHHDYYIIATS 141
Db  123 FSLGFEFRPGHEYYISAS 141

RESULT 43
US-08-240-124-4
; Sequence 4, Application US/08240124
; Patent No. 5516658
; GENERAL INFORMATION:
;  APPLICANT: BECKMANN, M. P.
;  APPLICANT: CERRETTI, DOUGLAS P.
;  TITLE OF INVENTION: CYTOKINE THAT BINDS THE CELL SURFACE
;   RECEPTOR HEK
;  NUMBER OF SEQUENCES: 4
;  CORRESPONDENCE ADDRESS:
;   ADDRESSEE: IMMUNEX CORPORATION
;   STREET: 51 UNIVERSITY STREET
;   CITY: SEATTLE
;   STATE: WASHINGTON
;   COUNTRY: USA
;   ZIP: 98101
;  COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: Apple Macintosh
;   OPERATING SYSTEM: Apple System 7.1
;   SOFTWARE: Microsoft Word for Apple, Version 5.1a
;  CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/240,124
;   FILING DATE:
;   CLASSIFICATION: 435
;  PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/161,132
;   FILING DATE: 03-DEC-1993
;  PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/114,426
;   FILING DATE: 30-AUG-1993
;  PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/109,745
;   FILING DATE: 20-AUG-1993
;  ATTORNEY/AGENT INFORMATION:
;   NAME: SEESE, KATHRYN A.
;   REGISTRATION NUMBER: 32,172
;   REFERENCE/DOCKET NUMBER: 2814-C
;  TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (206) 587-0430
;   TELEFAX: (206) 233-0644
;   TELEX: 756822
;
; INFORMATION FOR SEQ ID NO: 4:
;  SEQUENCE CHARACTERISTICS:
;   LENGTH: 201 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-240-124-4

Query Match       6.9%; Score 169.5; DB 1; Length 201;
Best Local Similarity   29.9%; Pred. No. 8e-07;
Matches  66; Conservative  18; Mismatches  82; Indels  55; Gaps  10;

Qy   24 LVSGLSL-EPVYWNSANKRFQAEGGYVLYPQIGDRLDLLCPRARPPGPHSSPNYEFYKLY 82
            ||:|:|  |||||| |:                 |:|  || :||  ||
Db   20 LRGGSSLRHVVWNSSNPRL-LRGDAVVELGLNDYLDIVCPHYEGGPHEGP--ETFALY 76

Qy   83 LVGGAQGRRCEAP-PAPNLLLTCDRPDLDLRFTIKFQEYSPNLWGHEFRSHHDYYIIATS 141
            :|   |  |  | ||||| |:|| :|||:||| || | |:|| ||||:||||:|:
Db   77 MVDWPGYESCQABGPRAYKRWVCSLPFGHVQFSEKIQRFTPFFSLGFEFLPGETYYISVP 136

Qy  142 DGTREGLESLQGGVCLTRGNKVLLRVGQSPRGGAVPRKFVSEMPMERDRGAAHSLEPGKE 201
            :|:
Db  137 --TPE-----SSGQCL-------------------RLQVSVCCKBERKSESAHPV----- 164
```

Left column

```
Qy  202 NLPGDPTSNATS--RGAEGPLPPPSMPAVAGAAGGLALLLL 240
            ||||  |||||
Db  165 ---GSPGESGTSGWRGGDTPSP-----------LCLLLL 189
```

RESULT 44
US-08-453-943-4
Sequence 4, Application US/08453943
; Patent No. 5738844
; GENERAL INFORMATION:
; APPLICANT: BECKMANN, M. P.
; APPLICANT: CERRETTI, DOUGLAS P.
; TITLE OF INVENTION: CYTOKINE THAT BINDS THE CELL SURFACE
; TITLE OF INVENTION: RECEPTOR HEK
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: IMMUNEX CORPORATION
; STREET: 51 UNIVERSITY STREET
; CITY: SEATTLE
; STATE: WASHINGTON
; COUNTRY: USA
; ZIP: 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Apple System 7.1
; SOFTWARE: Microsoft Word for Apple, Version 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/453,943
; FILING DATE: 30-MAY-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/240,124
; FILING DATE: 09-MAY-1994
; APPLICATION NUMBER: US 08/161,132
; FILING DATE: 03-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/114,426
; FILING DATE: 30-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/109,745
; FILING DATE: 20-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: SEESE, KATHRYN A.
; REGISTRATION NUMBER: 32,172
; REFERENCE/DOCKET NUMBER: 2814-C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 587-0430
; TELEFAX: (206) 233-0644
; TELEX: 756822
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 201 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-453-943-4

```
Query Match          29.9%;  Score 169.5;  DB 1;  Length 201;
Best Local Similarity  6.9%;
Matches  66;  Conservative  18;  Pred. No. 8e-07;  Mismatches 82;  Indels 55;  Gaps 10;

Qy  24 LVSGLSL-EPVYWNSANKRFQAEGGYVLYPQIGDRLLDLLCPRARPPGPHSSPNYEFYKLY 82
         |.|||||:|||||||||.||::.:| |:|||:||.||||||||||||||
Db  20 LRGGSSLRHVVYWNSSNPRL-LRGDAVELGLNDYLDIVCPHYEGPGPPEGP--ETFALY 76

Qy  83 LVGGAQGRRCEAP-PAPNLLLTCDRPDLDLRFTIKFQEYSPNLWGHEFRSHHDYYIIATS 141
         ||||.|:::|.|  |:.||:|:|||..|:  :.|:|:|||:||:
Db  77 MVDWPGYESCQAEGPRAYKRWVCSLPFGHVQFSEKIQRTPFSLGFEFLPGETYYYISVP 136

Qy 142 DGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKFVSEMPMERDRGAAHSLEFGKE 201
         |||  :|| :
Db 137 --TPS-----SSGQCL 164
```

Right column

```
Qy  202 NLPGDPTSNATS--RGAEGPLPPPSMPAVAGAAGGLALLLL 240
            ||||  |||||
Db  165 ---GSPGESGTSGWRGGDTPSP-----------LCLLLL 189
```

RESULT 45
US-09-057-121-4
Sequence 4, Application US/09057121
; Patent No. 5969110
; GENERAL INFORMATION:
; APPLICANT: BECKMANN, M. P.
; APPLICANT: CERRETTI, DOUGLAS P.
; TITLE OF INVENTION: CYTOKINE THAT BINDS THE CELL SURFACE
; TITLE OF INVENTION: RECEPTOR HEK
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: IMMUNEX CORPORATION
; STREET: 51 UNIVERSITY STREET
; CITY: SEATTLE
; STATE: WASHINGTON
; COUNTRY: USA
; ZIP: 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Apple System 7.1
; SOFTWARE: Microsoft Word for Apple, Version 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/057,121
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/240,124
; FILING DATE: 09-MAY-1994
; APPLICATION NUMBER: US 08/161,132
; FILING DATE: 03-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/114,426
; FILING DATE: 30-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/109,745
; FILING DATE: 20-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: SEESE, KATHRYN A.
; REGISTRATION NUMBER: 32,172
; REFERENCE/DOCKET NUMBER: 2814-C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 587-0430
; TELEFAX: (206) 233-0644
; TELEX: 756822
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 201 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-057-121-4

```
Query Match          29.9%;  Score 169.5;  DB 2;  Length 201;
Best Local Similarity  6.9%;
Matches  66;  Conservative  18;  Pred. No. 8e-07;  Mismatches 82;  Indels 55;  Gaps 10;

Qy  24 LVSGLSL-EPVYWNSANKRFQAEGGYVLYPQIGDRLLDLLCPRARPPGPHSSPNYEFYKLY 82
         |.|||||:|||||||||.||::.:| |:|||:||.||||||||||||||
Db  20 LRGGSSLRHVVYWNSSNPRL-LRGDAVELGLNDYLDIVCPHYEGPGPPEGP--ETFALY 76

Qy  83 LVGGAQGRRCEAP-PAPNLLLTCDRPDLDLRFTIKFQEYSPNLWGHEFRSHHDYYIIATS 141
         ||||.|:::|.|  |:.||:|:|||..|:  :.|:|:|||:||:
Db  77 MVDWPGYESCQAEGPRAYKRWVCSLPFGHVQFSEKIQRTPFSLGFEFLPGETYYYISVP 136

Qy 142 DGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKFVSEMPMERDRGAAHSLEFGKE 201
```

```
Db  137  --TPE-----SSGQCL------RLQVSVCCKERKSESAHPV-----  164
Qy  202  NLPGDPTSNATS--RGAEGPLPPPSMPAVAGAAGGLALLLL  240
Db  165  ---GSPGEGGTSGWRGGDTPSP--------------LCLLL  189
```

RESULT 46
US-09-358-734-4
; Sequence 4, Application US/09358734
; Patent No. 6274117
; GENERAL INFORMATION:
; APPLICANT: BECKMANN, M. P.
; APPLICANT: CERRETTI, DOUGLAS P.
; TITLE OF INVENTION: CYTOKINE THAT BINDS THE CELL SURFACE
; TITLE OF INVENTION: RECEPTOR HEK
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: IMMUNEX CORPORATION
;   STREET: 51 UNIVERSITY STREET
;   CITY: SEATTLE
;   STATE: WASHINGTON
;   COUNTRY: USA
;   ZIP: 98101
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: Apple Macintosh
;   OPERATING SYSTEM: Apple System 7.1
;   SOFTWARE: Microsoft Word for Apple, Version 5.1a
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/358,734
;   FILING DATE:
;   CLASSIFICATION:
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/240,124
;   FILING DATE:
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/114,426
;   FILING DATE: 30-AUG-1993
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/109,745
;   FILING DATE: 20-AUG-1993
; ATTORNEY/AGENT INFORMATION:
;   NAME: SEESE, KATHRYN A.
;   REGISTRATION NUMBER: 32,172
;   REFERENCE/DOCKET NUMBER: 2814-C
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (206) 587-0430
;   TELEFAX: (206) 233-0644
;   TELEX: 756822
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 201 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein

Query Match   6.9%;   Score 169.5;   DB 4;   Length 201;
Best Local Similarity   29.9%;   Pred. No. 8e-07;
Matches   66;   Conservative   18;   Mismatches   82;   Indels   55;   Gaps   10;

```
Qy   24  LIVSGLSL-EPVYWNSANRFQAEGGYVLYPQIGDRLDLLCPRARPPGPHSSSPNYEFYKLY   82
             ||||   |||||| || : |||  ||| :|  ::: ||||| ::|
Db   20  LRGGSSLRHVVYWNSSNPRL-LRGDAVELGLNDYLDIVCPHYEGPGPPGP--ETFALY   76

Qy   83  LVGGAQGRRCEAP-PAPNLLLTCDRPDLDLRFILKFQEYSPNLWGHERSHHDYYIIATS  141
             | |  |   |   ||:|| ||  :| | :||   ||  ||      :|:|| | :
Db   77  MVDWFGYESCQAEGPRAYXRWVCSLPFGHVQFSEKIQRTPFSLGFEFLPGETYYISVP  136

Qy  142  DGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKPVSEMPMERDRGAAHSLEPGKE  201
                                   :                            :
Db  137  --TPE-----SSGQCL-------RLQVSVCCKERKSESAHPV-----  164
```

```
Qy  202  NLPGDPTSNATS--RGAEGPLPPPSMPAVAGAAGGLALLLL  240
              :  || | :    |||| | ||  ||           |||||
Db  165  ---GSPGEGGTSGWRGGDTPSP-------------LCLLL  189
```

RESULT 47
US-09-214-631-8
; Sequence 8, Application US/09214631
; Patent No. 6413730
; GENERAL INFORMATION:
; APPLICANT: Holland, Sacha
; APPLICANT: Mbamalu, Geraldine
; APPLICANT: Pawson, Tony
; TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED
; TITLE OF INVENTION: BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
;   FILE REFERENCE: 11757.23USWO
; CURRENT APPLICATION NUMBER: US/09/214,631
; CURRENT APPLICATION NUMBER: 1999-03-12
; EARLIER APPLICATION NUMBER: PCT/CA97/00473
; EARLIER FILING DATE: 1997-07-04
; EARLIER APPLICATION NUMBER: 60/021,272
; EARLIER FILING DATE: 1996-07-05
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 8
; LENGTH: 201
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-214-631-8

Query Match   6.9%;   Score 169;   DB 4;   Length 201;
Best Local Similarity   29.8%;   Pred. No. 8.8e-07;
Matches   70;   Conservative   19;   Mismatches   84;   Indels   62;   Gaps   11;

```
Qy   17  LLLGVL------GLVSGLSL-EPVYWNSANRFQAEGGYVLYPQIGDRLDLLCPRARPP   68
            |::                 ||||     ||||| ||: |||  :||
Db    6  LIITVLNAAFLGSPLRGGSSLRHVVYWNSSNPR-SLRGDAVVELGLNDYLDIVCPHYEGP   64

Qy   69  GPHSSSPNYEFYKLYLVGGAQGRRCEAP-PAPNLLLTCDRPDLDLRFILKFQEYSPNLWGH  127
              :||:||:|| | |  | |  |   |   ||:|| ||  :|| :|: ||
Db   65  GPPEGP--ETFALYMVDWFGYESCQAEGPRGYKRWVCSLPFGHVQFSEKIQRTPFSLGF  122

Qy  128  EFRSHHDYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKPVSEMPME  187
              |:  ||| |  :                   :
Db  123  EFLPGETYYISVP--TPE-----SSGQCL--------RLQVSVCCKE  155

Qy  188  RDRGAAHSLEPGKENLPGDPTSNATS--RGAEGPLPPPSMPAVAGAAGGLALLLL  240
              :                :  || | :    |||| | ||  ||           |||||
Db  156  RKSESAHPV-----GSPGEGGTSGWRGGDTPSP-------------LCLLL  189
```

RESULT 48
US-09-214-631-6
; Sequence 6, Application US/09214631
; Patent No. 6413730
; GENERAL INFORMATION:
; APPLICANT: Holland, Sacha
; APPLICANT: Mbamalu, Geraldine
; APPLICANT: Pawson, Tony
; TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED
; TITLE OF INVENTION: BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR
; TITLE OF INVENTION: TYROSINE KINASES
; FILE REFERENCE: 11757.23USWO
; CURRENT APPLICATION NUMBER: US/09/214,631
; CURRENT APPLICATION NUMBER: 1999-03-12
; EARLIER APPLICATION NUMBER: PCT/CA97/00473
; EARLIER FILING DATE: 1997-07-04
; EARLIER APPLICATION NUMBER: 60/021,272
; EARLIER FILING DATE: 1996-07-05
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patentin Ver. 2.0

; SEQ ID NO 6
; LENGTH: 209
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-214-631-6

Qy   Query Match
       Best Local Similarity   35.9%;
       Matches   42;   Conservative   14;   Mismatches   52;   Indels   9;   Gaps   3;
                                                6.9%;   Score 168.5;   DB 4;   Length 209;

Qy   33 VYWNSANKFQAE--------GGYVLYPQIGDRLDLLCPRARPPGPHSSSPNYEFYKLYLVGGA 87
          ||| |||||||||:||||| |:||:||
Db   35 VYWNRSNPRFQVSAVGDGGGYTVEVSINDYLDIYCPHYGAPLP--PAERMERYILYMVNGE 93

Qy   88 QGRRCEAPPAPNLLLTCDRPDL----DLRFTIKFQEYSPNLWGHEFRSHHDYYIIATS 141
          |:|||| ||:|||:| :||||||||||:||||||||:|| |
Db   94 GHASCDHRORGFKRWECNRPAAPGGPLKFSEKFQLFTPFSLGFEFPPGHEYYYISAT 150

RESULT 49
US-08-442-248-4
; Sequence 4, Application US/0842248
; Patent No. 5759863
; GENERAL INFORMATION:
;   APPLICANT: Caras, Ingrid W.
;   APPLICANT: Winslow, John W.
;   TITLE OF INVENTION: AL-1 Neurotrophic Factor
;   NUMBER OF SEQUENCES: 18
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Genentech, Inc.
;   STREET: 460 Point San Bruno Blvd
;   CITY: South San Francisco
;   STATE: California
;   COUNTRY: USA
;   ZIP: 94080
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: patin (Genentech)
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/442,248
;   FILING DATE: 15-MAY-1995
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/330128
;   FILING DATE: 27-OCT-1994
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Torchia, Timothy E.
;   REGISTRATION NUMBER: 36,700
;   REFERENCE/DOCKET NUMBER: 920C4
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 415/225-8674
;   TELEFAX: 415/952-9881
;   TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 4:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 228 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
US-08-442-248-4

Qy   Query Match
       Best Local Similarity   28.8%;
       Matches   65;   Conservative   29;   Mismatches   81;   Indels   51;   Gaps   11;
                                                6.8%;   Score 167.5;   DB 1;   Length 228;   Pred. No. 1.4e-06;

Qy   33 VYWNSANKFQAEGGYVLYPQIGDRLDLLCPRARPPGPHSSSPNYEFYKLYLVG------- 85
          ||||||| ||||||:::||::|||||||||
Db   34 VYWNSSNPRFQ-RGDYHIDVCINDYLDVFCPHYEDSVPEDKT--ERYVLYMVNFDGYSAC 90

Qy   86 -----GAQGRRCEAPPAPNLLLTCDRPDLDLRFTIKFQEYSPNLWGHEFRSHHDYYIIAT 140
               :|: ||:|||:| :||||||||||:||||||||:||
Db   91 DHTSKGFKRWECNRPHSPN------------GPLKFSEKFQLFTPFSLGFEFPRGREYFYISS 141

OM protein - protein search, using sw model

```
                GenCore version 5.1.3
                Copyright (c) 1993 - 2003 Compugen Ltd.
```

Run on: February 11, 2003, 11:48:59 ; Search time 33.195 seconds
                                       (without alignments)
                                       1317.705 Million cell updates/sec

Title: US-10-021-121-2
Perfect score: 2450
Sequence: 1 MGPPHSGPGGVRVGALLLLG..........TTLLRQPASVEABAGQHGPL 455
Scoring table: BLOSUM62
                 Gapop 10.0 , Gapext 0.5
Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 50 summaries

Database :   PIR_73:*
                 1: pir1:*
                 2: pir2:*
                 3: pir3:*
                 4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB	ID	Description
1	628.5	25.7	336	2	I49766	hepatoma transmemb
2	623	25.4	346	2	S46993	elk ligand - human
3	620.5	25.3	333	2	I84743	hepatoma transmemb
4	604.5	24.7	345	2	I48780	Stra1/Epl2 protei
5	599.5	24.5	345	2	I58406	LERK-2 - rat
6	214.5	8.8	237	2	T19914	hypothetical prote
7	179	7.3	238	2	I38849	LERK-3 - human
8	176	7.2	209	2	A54984	ELF-1 protein prec
9	175.5	7.2	213	2	JE0322	ephrin-A2 - human
10	170.5	7.0	228	2	A57084	repulsive axon gui
11	169.5	6.9	201	2	I38850	LERK-4 - human
12	167.5	6.8	228	2	I58170	LERK-7 precursor -
13	166	6.8	205	2	A36377	B61 protein precur
14	159	6.5	680	2	S31216	collagen alpha 1(X
15	154.5	6.3	1049	2	CGB07S	collagen alpha 1(I
16	153.5	6.3	1670	1	CGHU3B	collagen alpha 3(I
17	151.5	6.2	301	2	B31219	collagen 2 - Caeno
18	149	6.1	325	2	T32248	hypothetical prote
19	149	6.1	569	2	S42886	collagen - silkwor
20	148.5	6.1	316	2	T20497	hypothetical prote
21	148	6.0	921	2	S42617	collagen alpha 1(I
22	146.5	6.0	1315	2	A56101	collagen alpha 1(X
23	146.5	6.0	1492	2	A40333	collagen alpha 1'(
24	146.5	6.0	1774	2	B56101	collagen alpha 1(X
25	146	6.0	675	2	S20819	collagen alpha 3(I
26	145.5	6.0	305	2	T20906	hypothetical prote
27	145	5.9	674	2	S13301	collagen alpha 1(I
28	145	5.9	931	2	S13580	collagen alpha 1(I
29	144.5	5.9	438	2	S53787	collagen alpha cha
30	144	5.9	1027	2	S28774	collagen alpha cha
31	143	5.8	1747	2	A54121	collagen alpha-4 c
32	142.5	5.8	635	2	A57131	collagen alpha 2(V
33	142.5	5.8	743	2	S23779	collagen alpha 1 (
34	142.5	5.8	1496	1	CGHU2V	collagen alpha 2 (V
35	142	5.8	614	2	T33149	hypothetical prote
36	142	5.8	744	2	S15435	collagen alpha 2(V
37	142	5.8	1029	1	S21369	collagen alpha 1(V
38	142	5.8	1763	2	S16366	collagen alpha 1(V
39	141.5	5.8	1466	1	CGHU7L	collagen alpha 1 (
40	141	5.8	319	2	T32250	hypothetical prote
41	141	5.8	744	1	A34246	collagen alpha 1(V
42	141	5.8	744	1	S23298	collagen alpha 1(V
43	140.5	5.7	305	2	T30165	hypothetical prote
44	140.5	5.7	304	2	T22482	hypothetical prote
45	140	5.7	680	1	CGHU1D	collagen alpha 1(X
46	139.5	5.7	210	2	B44984	collagen - nematod
47	139	5.7	940	2	JE0291	collagen alpha 1(X
48	139	5.7	1024	2	S18251	collagen alpha 2(V
49	138.5	5.7	469	2	A24450	collagen alpha 1(I
50	138.5	5.7	1419	2	A41182	collagen alpha 1(I
```

ALIGNMENTS

```
RESULT 1
I49766
hepatoma transmembrane kinase ligand - mouse
C;Species: Mus musculus (house mouse)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 05-Nov-1999
C;Accession: I49766
R;Bennett, B.D.; Zeigler, F.C.; Gu, Q.; Fendly, B.; Goddard, A.D.; Gillett, N.; Matthew
Proc. Natl. Acad. Sci. U.S.A. 92; 1866-1870, 1995
A;Title: Molecular cloning of a ligand for the EPH-related receptor protein-tyrosine ki
A;Reference number: I49766; MUID:95199254; PMID:7534404
A;Accession: I49766
A;Molecule type: mRNA
A;Residues: 1-336 <RES>
A;Cross-references: GB:L38847; NID:g769677; PIDN:AAC42052.1; PID:g769678
A;Genetics:
A;Gene: HTK

Query Match      25.7%;  Score 628.5;  DB 2;  Length 336;
Best Local Similarity   41.8%;  Pred. No. 3.1e-37;
Matches 137; Conservative 49; Mismatches 129; Indels 13; Gaps 5;

Qy    14  GALLLLGVLGLVSGLSLEPVYWNSANRKFQAEGGYVLYPQIGDRLDLLCPRARPGPHSS  73
              :|:|.:|| |.| | |:::|       | ::|||.: | .:|.|::|  |     ||
Db    17  GILMVLCRTAISRSIVLEPIYWNSSNSKFLPGQGLVLYPQIGDKLDIICPKV---DSKTV  73

Qy    74  PNYEFYKLYLVGGAQGRRCEAPAPNLLLTCDRPDLLDRFTIKFQEYSPNLWGHEFRSHH  133
              |.|::|  : || .|| ||  : .||:: || | ||.:| |:| |  |:: |:.
Db    74  GQYEYKVYMVDKDQADRCTIKKENTPLLNCARPDQDVKFTIKFQEFSPNLWGLEFQKNK  133

Qy    134 DYYIIATSDGTRGLESLQGGVCLTRGMKVLLRVGQ---SPRGGAVPRKPVSEMPER-DR  190
              | ||:|:|| .:|::: |:|:|:  .|::|  .|     | ::|.  .:  |::  .
Db    134 DYYIISTENGSLEGLDNQEGGVCQTRAMKILMKVGQDASSAGSARNHGPTRRPELEAGTN  193

Qy    191 GAAHSLEPGKENLPGDPTSNAISRGAEGPLPPSMPAVAGAAGGLALLLLGYVAGAGAMC  250
              |   .|:|| :|| .|::: : ||| || | | :|.  |.| || |:::|  .|
Db    194 GRSSTTSPFVKPNPGSSTDGNSAGHSGNNLLGSEVALPAGIASGCIIFIVIILVVLLL  253

Qy    251 WRRRRAKPSESRHPGPGSFGRGGSLGLGGCGGGMGPRRAEPGELGIALRGGGAADPPCPH  310
              |:::: | :: :: |::|::   ||.||.|: |: ||  ||::| || ||..|  ||
Db    254 KYRRRHRCHSPQHTTTLSLSTLATPKRGGNN----NGSSEPSDVIIPLR---TADSVFCPH  306

Qy    311 YEKVSGDYGHPVIVQDGPPQSPPNIYY  338
              |||:|||:|||||||  |  || |||
Db    307 YEKVSGDYGHPVIVQEMPPQSPANIYY  334
```

RESULT 2
S46693
elk ligand - human
C;Species: Homo sapiens (man)
C;Date: 15-Jul-1995 #sequence_revision 10-Nov-1995 #text_change 28-May-1999
C;Accession: S46693
R;Beckmann, M.P.; Cerretti, D.P.; Baum, P.; vanden Bos, T.; James, L.; Farrah, T.; Kozlo
EMBO J. 13, 3757-3762, 1994
A;Title: Molecular characterization of a family of ligands for eph-related tyrosine kina
A;Reference number: S46693; MUID:94349923; PMID:8070404
A;Accession: S46693
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-346 <BEC>
A;Cross-references: GB:U09304; NID:g538366; PIDN:AAA53093.1; PID:g538367

```
Query Match         25.4%;  Score 623;  DB 2;  Length 346;
Best Local Similarity   39.2%;  Pred. No. 7.7e-37;
Matches 143;  Conservative 48;  Mismatches 116;  Indels 58;  Gaps 9;

Qy    8 PGGVRVGALLLLGVLGLVSGL------SLEPVYWNSANKRFQAEGGYVLYPQIGDRLDLL 61
Db    4 CPRABAGRP-----YEYYKLYLVRPEQAAACSTVLDPNVLVTCNRPEQEIRFTIKFQEFS 63

Qy   62 CPRARPPGPHSSPNYEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDLDLRFTIKFQEYS 121
Db   64 YEYYKLYLVRPEQAAACSTVLDPNVLVTCNRPEQEIRFTIKFQEFS 118

Qy  122 PNLWGHEFRSHDYYIIATSDGTREGLBSLQGGVCLITRGMKVLLRVGQSPRGGAVPRKPV 181
Db  119 PNYMGLEFKKHHDYYITSTSNGSLEGLENREGGVCRIRTMKIIMKVSGQDPNAVTPEQLTT 178

Qy  182 SEMPMERDRGAAHSLE-PGKENLPGDFTSNATSRGAEGPLPPPSMPAVAGAAGGLA---- 236
Db  179 SRPSKEADNTVKWATQAPGSRGSLGDSDGKHETVNQEEKSGP------GASGSGSGDPD 231

Qy  237 ------LILLLGVAGAGGA-----------MCWRRRRAKPSESSRHPQPGSFGRGGSLGL 277
Db  232 GFFNSKVALPAAVGAGCVIFLLIIIFLTVLLLKLRKHRKHTQQ-------RAAALSL 282

Qy  278 ----GGGGGMGPREAEPGELGIALRGGGAADPPFCPHYEKYSGDYGHPVYIVQODGPPQSP 333
Db  283 STLASPKGSGSGTAGTBPSDIIIPLR---TTENNYCPHYEKVSGDYGHPVYIVQEMPPQSP 339

Qy  334 PNIIY 338
Db  340 ANIYY 344
```

RESULT 3
I84743
hepatoma transmembrane kinase ligand - human
C;Species: Homo sapiens (man)
C;Date: 24-May-1996 #sequence_revision 24-May-1996 #text_change 05-Nov-1999
C;Accession: I84743
R;Bennett, B.D.; Zeigler, F.C.; Gu, Q.; Fendly, B.; Goddard, A.D.; Gillett, N.; Matthews
Proc. Natl. Acad. Sci. U.S.A. 92, 1866-1870, 1995
A;Title: Molecular cloning of a ligand for the EPH-related receptor protein-tyrosine kin
A;Reference number: I84743; MUID:95192254; PMID:7534404
A;Accession: I84743
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-333 <RES>
A;Cross-references: GB:L38734; NID:g769675; PIDN:AAC41752.1; PID:g769676
A;Gene: GDB:EPLG5; LERK5
A;Cross-references: GDB:433338; OMIM:600527
C;Genetics:
A;Map position: 13q33-13q33

```
Query Match         25.3%;  Score 620.5;  DB 2;  Length 333;
Best Local Similarity   40.5%;  Pred. No. 1.1e-36;
Matches 133;  Conservative 52;  Mismatches 130;  Indels 13;  Gaps 5;
```

```
Qy   14 GALLLLGVLGLVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDLLCPRARPPGPHSS 73
Db   14 GVLMVLCRTAISKSIVLEPIYWNSSNSKFLPGQGLIVLYPQIGDKLDIICPKV--DSKTV 70

Qy   74 PNYEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDLDLRFTIKFQEYSPNLWGHEFRSHH 133
Db   71 GQVEYYKIYKVYMDKDQADRCITKKENTPLLNCAKPDQDIKFTIKFQEFSPNLWGLEFQKNK 130

Qy  134 DYYIIATSDGTREGLBSLQGGVCLITRGMKVLLRVGQ--SPRGGAVPRKPVSEMPVER-DR 190
Db  131 DYYIISTSNGSLEGLDNQEGGVCQTRAMKILMKVGQDASAGSTRNKDPTRRPELEAGTN 190

Qy  191 GAAHSLEPGKENLPGDFTSNATSRGAEGPLPPPSMPAVAGAAGGLALLLLGVAGAGGAMC 250
Db  191 GRSSTTSPFVKPNPGSSTDGNSAGHSGNNILGSEVALFAGIASGCIIFIVIITLVVLLL 250

Qy  251 WRRRRRAKPSESRHPGPGSFGRGGSLGLGGGGGMGPREAEPGELGIALRGGGAADPPFCPH 310
Db  251 KYRRRHRKHSPQHTTTLSLSTLATPKRSGNN----NGSEPSDIIIPLR---TADSVFCPH 303

Qy  311 YEKVSGDYGHPVYIVQDGPPQSPPNIIY 338
Db  304 YEKVSGDYGHPVYIVQEMPPQSPANIYY 331
```

RESULT 4
I48780
Stral/Eplg2 protein - mouse
C;Species: Mus musculus (house mouse)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 05-Nov-1999
C;Accession: I48780; A55507; A550062; S52670
R;Bouillet, P.; Oulad-Abdelghani, M.; Vicaire, S.; Garnier, J.M.; Schuhbaur, B.; Dolle,
Dev. Biol. 170, 420-433, 1995
A;Title: Efficient cloning of cDNAs of retinoic acid-responsive genes in P19 embryonal
A;Reference number: I48780; MUID:95377533; PMID:7649373
A;Accession: I48780
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-345 <RES>
A;Cross-references: EMBL:Z48781; NID:g747858; PIDN:CAA88695.1; PID:g747859
R;Fletcher, F.A.; Renshaw, B.; Hollingsworth, T.; Baum, P.; Lyman, S.D.; Jenkins, N.A.;
Genomics 24, 127-132, 1994
A;Title: Genomic organization and chromosomal localization of mouse Eplg2, a gene encod
A;Reference number: A55507; MUID:95203867; PMID:7896266
A;Accession: A55507
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-345 <FLE>
A;Cross-references: GB:U07598
R;Shao, H.; Lou, L.; Pandey, A.; Pasquale, E.B.; Dixit, V.M.
J. Biol. Chem. 269, 26606-26609, 1994
A;Title: cDNA cloning and characterization of a ligand for the Cek5 receptor protein-ty
A;Reference number: A55062; MUID:95014510; PMID:7929389
A;Accession: A55062
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-89, 'T', 91-345 <SHA>
A;Cross-references: GB:U12983; NID:g575928; PIDN:AAA52231.1; PID:g575929
C;Genetics:
A;Gene: EPLG2

```
Query Match         24.7%;  Score 604.5;  DB 2;  Length 345;
Best Local Similarity   37.9%;  Pred. No. 1.6e-35;
Matches 136;  Conservative 51;  Mismatches 107;  Indels 65;  Gaps 10;

Qy   15 ALLLLGVLGLVSGL--SLEPVYWNSANKRFQAEGGYVLYPQIGDRLDLLCPRARPPGPHS 72
Db   15 AMVVLTLCRLATPLAKNLEPVSWSSLNPKFLSGKGLVIYPKIGDKLDIICPRAEAGRP-- 72

Qy   73 SPNYEFYKLYLVGGAQGRRCEAPAPNLLLTCDRPDLDLRFTIKFQEYSPNLWGHEFRSH 132
Db   73 ---YEYYKLYIVRPEQAAACSTVLDPNVLVTCNRPHQEIRFTIKFQEFSPNYMGLEFKKY 129
```

```
Qy    133  HDYYIIATSDGTREGLESLQGGVCLITRGMKVLLRVGQSPRGGAVPRKPVSEMPMERDRGA 192
               |:||| |:|||   |||||||||| |:||||||||||:  |::|||  : |
Db    130  HDYYITSTSNGSLEGLENREGGVCRTRTMKIVMKVGQDP-NAVTPEQLTTSRPSKESDNT 188

Qy    193  AHSLEPGKENLPGDPTSNATSRGAEGP-------------LPPPSMPAVAGAAGG------LA 236
               |:|::                                  ::||| |  |||
Db    189  VKT-------------------------ATQAPGRGSGQGDSDGKHETVNQEKSGPGAGGSGDSDSFFNSK 236

Qy    237  LLLLGVAGAGA---------------------------MCWRRRRAKPSESRHPGPGSFGRGGSLGL----GG 279
                                                       |:|  :   :|
Db    237  VALFAAVGAGCVIFLLIIIFLTVLLLKLRHRHRHTQQ-------------RAAALSLSTLASP 287

Qy    280  GGGGMGPREAEPGELGIALRGGGAADPPFCPHYEKVSGDYGHPYYIVQDGPPQSPPNIIY 338
               |: :  | | | |:| ||||
Db    288  KGGSGTAGTEPSDIIIPLR---TTENNYCHYEKVSGDYGHPVVIVQEMPPQSPANIIY 343


RESULT 5
I58406
LERK-2 - rat
C;Species: Rattus norvegicus (Norway rat)
A;Accession: I58406
R;Fletcher, F.A.; Carpenter, M.; Shilling, H.; Baum, P.; Ziegler, S.; Gimpel, S.; Hollin
Oncogene 9, 3241-3248, 1994
A;Title: LERK-2, a ligand for the receptor tyrosine kinase ELK, is evolutionarily conser
A;Reference number: I58406; MUID:95022634; PMID:7936648
A;Accession: I58406
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-345 <RES>
A;Cross-references: EMBL:U07560; NID:g563118; PIDN:AAA53092.1; PID:g563119
A;Genetics:
A;Gene: Eplg2

Query Match             24.5%;  Score 599.5;  DB 2;  Length 345;
Best Local Similarity   37.6%;  Pred. No. 3.5e-35;
Matches 135;  Conservative 52;  Mismatches 107;  Indels 65;  Gaps 10;

Qy     15  ALLLLGVLGLVSGL--SLEPVYWNSANKRFQAEGGYVLYFPQIGDRLDLLCPRARPPGPHS 72
               |:|| ||:|:||     ||||| :|||||||||||||||   |:|:||:|||| |::
Db     15  AMVVLTLCRLATPLAKNLEPVSWSSLNPKFLSGKGLVIPKIGDKLDIICPRABAGRP-- 72

Qy     73  SPNYEFYKLYLVGGAQGRRCEAPPAAPNLLTCDRPDLDLRFTIKFQEYSPNLWGHEFRSH 132
               ||| |:|| :||:||||||||||||||||:|| | ::: ||: |||||||| |||| :
Db     73  ---YEYYKYLVRPEOAAACSTVLDPNVLVTCNKPQQEIRFTIKFQEFSPNYMGLEFKKY 129

Qy    133  HDYYIIATSDGTREGLESLQGGVCLITRGMKVLLRVGQSPRGGAVPRKPVSEMPMERDRGA 192
               |:||| |:|||   |||||||||| |:||||||||||:  |::|||  : |
Db    130  HDYYITSTSNGSLEGLENREGGVCRTRTMKIVMKVGQDD-NAVTPEQLTTSRSKESDNT 188

Qy    193  AHSLEPGKENLPGDPTSNATSRGAEGP-------------LPPPSMPAVAGAAGG------LA 236
               |:|::                                  ::||| |  |||
Db    189  VKT-------------------------ATQAPGRGSGQGDSDGKHETVNQEKSGPGAGGSGDTDSFPNSK 236

Qy    237  LLLLGVAGAGA---------------------------MCWRRRRAKPSESRHPGPGSFGRGGSLGL----GG 279
                                                       |:|  :   :|
Db    237  VALFAAVGAGCVIFLLIIFLTVLLLKLRHRHRHTQQ-------------RAAALSLSTLASP 287

Qy    280  GGGGMGPREAEPGELGIALRGGGAADPPFCPHYEKVSGDYGHPYYIVQDGPPQSPPNIIY 338
               |: :  | | | |:| ||||
Db    288  KGDSGTAGTEPSDIIIPLR---TTENNYCHYEKVSGDYGHPVVIVQEMPPQSPANIIY 343


RESULT 6
T19914
hypothetical protein C43F9.8 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T19914
R;Mortimore, B.
submitted to the EMBL Data Library, November 1996
A;Reference number: Z19195
```

```
A;Accession: T19914
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-237 <WIL>
A;Cross-references: EMBL:Z82262; PIDN:CAB54195.1; GSPDB:GN000022; CESP:C43F9.8
A;Experimental source: clone C43F9
C;Genetics:
A;Gene: CESP:C43F9.8
A;Map position: 4
A;Introns: 32/2; 96/3; 214/1

Query Match             8.8%;  Score 214.5;  DB 2;  Length 237;
Best Local Similarity   25.7%;  Pred. No. 4e-08;
Matches 53;  Conservative 41;  Mismatches 83;  Indels 29;  Gaps 6;

Qy     11  VRVGALLLLGVLGLVS-GLSLEPVYWNSANKRFQAEG-GYVLYFPQIGDRLDLLICPRARPP 68
               |:|| |||:||:|||  |::||||||||||||||||||||   ||||:|:|||  |:|||
Db      1  MQIATFILLSLPFFIGWARKIPDINWISSNPIFDVSNTDHVISVHIGDRVSIRCPKSDET 60

Qy     69  GPHSSPNYEFYKLYLVGGAQGRRCEAPPAAPNLLLTCDRPDLDLRFTIKFQEYSPNLWGHE 128
               |||::||| |:||||:  :|:||||||||||||||||||| |: |||||:|:||||
Db     61  G-----KYEYSIYMVSDEBYDHCFL-SKPRLVGACDNQTINASINIVFRSFTPTGGPE 114

Qy    129  FRSHHDYYIIA-------------------TSDGTREGLESLQGGVCLITRGMKVLLRVGQSPRGGAVP 169
               |:|:|||||                      |||||||||||||||| |||||||||||||
Db    115  FQPGKNYFLISKSEVDALIIYETANQIPPGTSDGTLEGIDRKKDGLCTAKQMIKFEVGQ 174

Qy    170  SPRGGAVPRKPVSEMPMERDRGAAHS 195
               |:|::||| || :
Db    175  DRGIENPX--FAARTLKKDRDABHS 198


RESULT 7
I38849
LERK-3 - human
C;Species: Homo sapiens (man)
C;Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 29-Sep-1999
C;Accession: I38849
R;Kozlosky, C.J.; Maraskovsky, E.; McGrew, J.T.; VandenBos, T.; Teepe, M.; Lyman, S.D.;
Oncogene 10, 299-306, 1995
A;Title: Ligands for the receptor tyrosine kinases hek and elk: isolation of cDNAs enco
A;Reference number: I38849; MUID:95140419; PMID:7836529
A;Accession: I38849
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-238 <RES>
A;Cross-references: EMBL:U14187; NID:g642832; PIDN:AAC50078.1; PID:g642833
C;Genetics:
A;Gene: GDB:EPLG3
A;Cross-references: GDB:438336; OMIM:601381
A;Map position: 1q21-1q22
C;Superfamily: axon guidance signal protein

Query Match             7.3%;  Score 179;  DB 2;  Length 238;
Best Local Similarity   28.4%;  Pred. No. 1.3e-05;
Matches 65;  Conservative 24;  Mismatches 80;  Indels 60;  Gaps 12;

Qy      7  GPGGVRVGALLLLGVLGLVSGLSLEPVYWNSANKRFQAEGGYVLYFPQIGDRLDLLCP--R 64
               |||||||||||||||||||||||||||||||||||   |:|||||  |:||:|||||
Db      1  GPHSSPNYEFYKLYLVGGAQGRRCEAPPAAPNLLLTCDRPDLDLRFTIKFQEYSPNLWGHE 60

Qy     65  ARPPGPHSSP-----NYEFYKLYLVGGAQGRRCEAPPAAPNLLLTCDRPDL---DLRFTIKF 117
               |:||| |                :|||||||||||||||||| |||:|||:    |||||
Db     61  G-----KYEYSIYMVSDEBYDHCFL-SKPRLVGACDNQTINASINIVFRSFTPTGGPE 114

Qy    118  QEYSPNLWGHEFRSHHDYYIIATSDGTREGLESLQGGVCLITRGMKVLLRVGQSPRGGAVP 177
               |:||||     |:|:|||||| |||||||||||||||||||||||||||||||||
Db    115  FQPGKNYFLISKSEVDALIIYETANQIPPGTSDGTLEGIDRKKDGLCTAKQMIKFEVGQ 174

Qy    178  RKPVSEMP----------MERDRGAAHSLE------PGKENLP 204
               || |:  |          |:|:||| ||:
Db    175  EKPVPTLPQFTMGPNVKINVLEDFBGENPQVPKLEKSISGTSPKREHLP 223
```

RESULT 8
A54984
ELF-1 protein precursor - mouse
N;Alternate names: Cek7 ligand
N;Species: Mus musculus (house mouse)
C;Accession: A54984
C;Date: 11-Nov-1994 #sequence_revision 11-Nov-1994 #text_change 29-Sep-1999
C;Molecule type: mRNA
R;Cheng, H.J.; Flanagan, J.G.
Cell 79, 157-168, 1994
A;Accession: A54984; A55873
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-209 <CHE>
A;Cross-references: GB;U14941; NID:g558836; PIDN:AAA53636.1; PID:g558837
A;Title: Identification and cloning of ELF-1, a developmentally expressed ligand for the
R;Shao, H.; Lou, L.; Pandey, A.; Verderame, M.F.; Siever, D.A.; Dixit, V.M.
J. Biol. Chem. 270, 3467-3470, 1995
A;Title: cDNA cloning and characterization of a Cek7 receptor protein-tyrosine kinase li
A;Accession: A55873
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-209 <SHA>
A;Cross-references: GB;U14752; NID:g681886; PIDN:AAA68520.1; PID:g681887
A;Superfamily: axon guidance signal protein
C;Keywords: lipoprotein; membrane protein

Query Match   29.3%; Score 176; DB 2; Length 209;
Best Local Similarity   29.3%;
Matches 58; Conservative 19; Mismatches 69; Indels 52; Gaps 7;

```
Qy   33 VYVNSANKRFQAE-------GGYVLYPQIGDRLDLLCPRARPPGPHSSPNYEFYKLYLVGGA 87
Db   35 VYWRSNPRFQVSAVGDGGGYTVEVSINDYLDIYCPHYGAPLD-PAERMEHYVLYMVNGE 93

Qy   88 QGRRCEAPPAPNLLLTCDRPDL---DLRFTIKFQEYSPNLWGHEFRSHHDYYIIATSDGT 144
Db   94 GHASCDHRQGFKRWECNRPAAPGGPLKFSEKFQLFTPFSLGFEFRPGHEYYYISATP-- 151

Qy  145 REGLESLQGGVCLTRGMKYLLLRVQQSPRGGAVPRKPVSEMPMERDRGAAHSLEPGKENLP 204
Db  152 ----PNLVDRPCLR-LKYTVR------------------------PTNETLY 174

Qy  205 GDP------TSNATSRGAEG 218
Db  175 EAPEPIFTSNSSCSGLGG 192
```

RESULT 9
JE0322
ephrin-A2 - human
C;Species: Homo sapiens (man)
C;Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 21-Jul-2000
C;Accession: JE0322
R;Aasheim, H.; Pedeutour, F.; Grosgeorge, J.; Logtenberg, T.
Biochem. Biophys. Res. Commun. 252, 378-382, 1998
A;Title: Cloning, chromosomal mapping, and tissue expression of the gene encoding the huma
A;Reference number: JE0322; MUID:99045414; PMID:9826538
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-213 <AAS>
A;Cross-references: GB;AJ007292; NID:g3688367; PIDN:CAA07435.1; PID:g3688368
A;Superfamily: axon guidance signal protein

Query Match   36.8%; Score 175.5; DB 2; Length 213;
Best Local Similarity   36.8%; Pred. No. 2e-05;
Matches 43; Conservative 14; Mismatches 51; Indels 9; Gaps 3;

```
Qy   33 VYVNSANKRFQA------EGGYVLVPQIGDRLDLLCPRARPPGPHSSPNYEFYKLYLVGGA 87
Db   39 VYWNRSNPRFHAGAGDGDGGYTVEVSINDYLDIYCPHYGAPLP-PAERMEHYVLYMVNGE 97

Qy   88 QGRRCEAPPAPNLLLTCDRPDL---DLRFTIKFQEYSPNLWGHEFRSHHDYYIIATS 141
Db   98 GHASCDHRQGFKRWECNRPAAPGGPLKFSEKFQLFTPFSLGFEFRPGHEYYYISAT 154
```

RESULT 10
A57084
repulsive axon guidance signal protein RAGS precursor - chicken
C;Species: Gallus gallus (chicken)
C;Date: 03-Nov-1995 #sequence_revision 03-Nov-1995 #text_change 20-Jun-2000
C;Accession: A57084
R;Drescher, U.; Kremoser, C.; Handwerker, C.; Loeschinger, J.; Noda, M.; Bonhoeffer, F.
Cell 82, 359-370, 1995
A;Title: In vitro guidance of retinal ganglion cell axons by RAGS, a 25 kDa tectal prot
A;Accession: A57084; MUID:95360980; PMID:7634326
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-228 <DRE>
A;Cross-references: GB;X90377; NID:g1061113; PIDN:CAA62027.1; PID:g984118
A;Superfamily: axon guidance signal protein
C;Keywords: glycoprotein; membrane protein; phosphatidylinositol linkage
F;1-20/Domain: signal sequence #status predicted <SIG>

Query Match   28.9%; Score 170.5; DB 2; Length 228;
Best Local Similarity   28.9%; Pred. No. 4.9e-05;
Matches 73; Conservative 30; Mismatches 91; Indels 59; Gaps 13;

```
Qy   16 LLLLGVLGL-VSGLSLEP--------------VYWNSANKRFQAEGGYVLYPQIGDRLDLLCPR 64
Db    6 MLLLAVAALWVCVRGQEPGRKAVADRYAVYWNSTNPRFQ-QGDYHYDVCINDYLDVFCPH 64

Qy   65 ARPPGPHSSPNYEFYKLYLVG-------------GAQGRRCEAPPAPNLLLTCDRPDLDLR 112
Db   65 YEDSVPEDKI-ERYVLYMVNFDGYSSCDHISKGFKRWECNRPHSPN------------GPLK 113

Qy  113 FTIKFQEYSPNLWGHEFRSHHDYYIIATS---DGTREGLESLQGGVCLTRGMKVLLRVQQ 169
Db  114 FSEKFQLFTPFSLGFEFRPGREYFYISSAIPDNGRRS--------CLK-LKVFVR---- 159

Qy  170 SPRGGAVPRKPVSEMPMERDRGAAHSLEPGKENLPGDPTSNATSRGAEGPLPPPSMP--A 227
Db  160 -PANSCMKTIGVHDRVPDVDNDKVENSLEPADDTV--RESAEPSRG-ENAAQTPRIPIRL 214

Qy  228 VAGAAGGLAILLL 240
Db  215 LATLILFLIAMLLI 227
```

RESULT 11
I38850
LERK-4 - human
C;Species: Homo sapiens (man)
C;Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 29-Sep-1999
C;Accession: I38850
R;Kozlosky, C.J.; Maraskovsky, E.; McGrew, J.T.; VandenBos, T.; Lyman, S.D.;
Oncogene 10, 299-306, 1995
A;Title: Ligands for the receptor tyrosine kinases hek and elk: isolation of cDNAs enc
A;Reference number: I38850; MUID:95140419; PMID:7838529
A;Accession: I38850
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-201 <RSS>
A;Cross-references: GB;EPLG4
A;Gene: GDB;EPLG4
A;Genetics:
A;Cross-references: GDB;438337; OMIM:601380
A;Cross-references: EMBL;U14188; NID:g642834; PIDN:AAC50079.1; PID:g642835
A;Map position: 1q21-1q22
A;Superfamily: axon guidance signal protein

Query Match   6.9%; Score 169.5; DB 2; Length 201;

```
Best Local Similarity 29.9%; Pred. No. 5e-05;
Matches     66; Conservative     18; Mismatches     82; Indels     55; Gaps     10;

Qy   24 LVSGVGLSL-EPVYWNSANKRFQAEGGYVLYPQIGDRLDLLCPRARPPGPHSSPNYEFYKLY 82
           |||||||:  .||:|||...|:|...||||||||||||||||||||:|||||||
Db   20 LRGGSLRHVVYWNSNPRL-LRGDAVVELGLNDYLDIVCPHYEGPGPPEGP--ETFALY 76

Qy   83 LVGGAQGRRCEAP-PAPNLLLTCDRPDLDLRFTIKFQEYSPNLWGHEFRSHHDYYIIATS 141
           |:|:.|| ||.||.|||||||||||||||||||||||||:|||||||||||||||
Db   77 MVDWPGYESCQAEGPRAYKEWVCSLPFGHVYQFSEKIQRFTPFSLGFEFLPGETYYYISVP 136

Qy  142 DGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKPVSEMPMERDRGAAHSLEPGKE 201
           |.|||||||||||||||||||||||||||||||||||||||||:..:||||:|.:|:
Db  137 --TPE-----SSGQCLI-----RLQVSVCCKERKSESAHPV----- 164

Qy  202 NLPGDPTSNATS--RGAEGPLPPPSMPAVAGAAGGLALLLL 240
           ||||||.||||||...|:|||
Db  165 ---GSPGEGSGTSGWRGGDTPSP--------------LCLLL 189
```

RESULT 12
I58170
LERK-7 precursor - human
N;Alternate names: AL-1
C;Species: Homo sapiens (man)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 29-Sep-1999
R;Kozlosky, C.J.; VandenBos, T.; Park, L.S.; Cerretti, D.P.; Carpenter, M.K.
R;Reference number: G08477
submitted to the EMBL Data Library, May 1995
A;Accession: I58170; G01812
R;Winslow, J.W.; Moran, P.; Valverde, J.; Shih, A.; Yuan, J.Q.; Wong, S.C.; Tsai, S.P.;
Neuron 14, 973-981, 1995
R;Title: Cloning of AL-1, a ligand for an Eph-related tyrosine kinase receptor involved
A;Reference number: I58170; MUID:95267434; PMID:7748564
A;Accession: I58170
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-228 <RES>
A;Cross-references: EMBL:U26403; NID:g1019430; PIDN:AAB60377.1; PID:g1019431
A;Cross-references: GB:S77167; NID:g914184; PID:g914185
A;Genetics:
A;Gene: GDB:EPLG7; AP1; LERK7
A;Cross-references: GDB:568757; OMIM:601535
A;Map position: 13q33-13q33
C;Superfamily: axon guidance signal protein

```
Query Match
Best Local Similarity  6.8%; Score 167.5; DB 2; Length 228;
Matches     65; Conservative     29; Mismatches     81; Indels     51; Gaps     11;

Qy   33 VYWNSANKRFQAEGGYVLYPQIGDRLDLLCPRARPPGPHSSPNYEFYKLYLVG------- 85
           ||||:|.|:|    |.|.|:||
Db   34 VYWNSSNPRFQ-RGDHIDVCINDYLDVFCPHYEDSVPEDKT--ERYVLVMVNFDGYSAC 90

Qy   86 -----GAQGRRCEAPPAPNLLLTCDRPDLDLRFTIKFQEYSPNLWGHEHRSHHDYYIIAT 140
               |.:|||  || :||||||||||||||||:|.|||||||||
Db   91 DHTSKGFKWENRPHSPN------------GPLKFSEKFQLFTPFSLGFEFRPGREYFYISS 141

Qy  141 S--DGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKPVSEMPMERDRGAAHSLE 197
           .   |.|:    |.|||||||||||||
Db  142 AIPDNGRRS------CLK--LKVFVR----PTNSCMKTIGVHDRVPDVDNDKVENSLE 186

Qy  198 PGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAGGLALLLLGVA 243
           |..:|||||||||||:|||||||||||||||||||:|
Db  187 PADDTV---HESAEFSRG-ENAAQTPIFSRL-----LAILLFLIA 223
```

RESULT 13
A36377
B61 protein precursor - human
C;Species: Homo sapiens (man)
C;Date: 28-Mar-1991 #sequence_revision 28-Mar-1991 #text_change 29-Sep-1999
A;Accession: A36377
R;Holzman, L.B.; Marks, R.M.; Dixit, V.M.
Mol. Cell. Biol. 10, 5830-5838, 1990
A;Title: A novel immediate-early response gene of endothelium is induced by cytokines an
A;Reference number: A36377; MUID:91042512; PMID:2233719
A;Accession: A36377
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-205 <HOL>
A;Cross-references: GB:M57730; GB:M37476; NID:g179320; PID:AAA58388.1; PID:g179321
C;Superfamily: axon guidance signal protein

```
Query Match
Best Local Similarity  6.8%; Score 166; DB 2; Length 205;
Matches     46; Conservative     31; Mismatches     74; Indels     16; Gaps     5;

Qy   18 LLGVLGLVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDLLCPRARPPGPHSSPN--75
           |||:  .:  |.||||:|||...:|.||    |.||:|:||||||||||
Db    8 LLGLCCSLAAADRHTVFWNSSNPKFRNE-DYTIHVQLNDYVDIICPHYE---DHSVADAA 63

Qy   76 YEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDL---DLRFTIKFQEYSPNLWGHEFRSH 132
           :|.|    .|:.||||  || :||||||||||||||||:|||||||
Db   64 MEQYILVLVEHEEYQLCQPQSKDCVRWQCNRSAKGPEKLSEKFQRFTPFTLGKEFKEG 123

Qy  133 HDYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRK 179
           |.|||: .  ||||||||||||||||||||:||||:|.||||
Db  124 HSYYIYISKPIHQHEDR------CLRLKVTVSGKITHSPQAHVNPQE 163
```

RESULT 14
S31216
collagen alpha 1(X) chain precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 13-Aug-1999
A;Accession: S31216; S28807; S22215; S30127; T48299; S26397; S31830
R;Kong, R.Y.C.; Kwan, K.M.; Lau, E.T.; Thomas, J.T.; Boot-Handford, R.P.; Grant, M.E.; C
Eur. J. Biochem. 213, 99-111, 1993
A;Title: Intron-exon structure, alternative use of promoter and expression of the mouse
A;Reference number: S31216; MUID:93287550; PMID:8477738
A;Accession: S31216
A;Molecule type: DNA
A;Residues: 1-680 <KON>
A;Cross-references: EMBL:Z21610; NID:g49793; PID:g49793; PIDN:CAA79736.1; PID:g49794
R;Elima, K.; Eerola, I.; Rosati, R.; Metsaeranta, M.; Garofalo, S.; Peraelae, M.; de Cro
Biochem. J. 289, 247-253, 1993
A;Title: Specific hybridization probes for mouse alpha-2(IX) and alpha-1(X) collagen mRN
A;Reference number: S22215; MUID:92182017; PMID:1543751
A;Accession: S22215
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 385-450, 'K', 452-627 <ELA>
A;Cross-references: EMBL:X63013; NID:g49795; PIDN:CAA44741.1; PID:g49796
R;Apte, S.S.; Olsen, B.R.
Matrix 13, 165-179, 1993
A;Title: Characterization of the mouse type X collagen gene.
A;Reference number: S30127; MUID:93261348; PMID:8492743
A;Accession: S30127
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-12, 'F', 14-26, 'S', 28-247, 'L', 249-285, 'A', 287-305, 'F', 307-416, 'S', 418-499, 'I
R;Apte, S.S.; Seldin, M.F.; Hayashi, M.; Olsen, B.R.
Eur. J. Biochem. 206, 217-224, 1992

RESULT 13
A36377

A;Title: Cloning of the human and mouse type X collagen genes and mapping of the mouse t
A;Reference number: I48299; MUID:92267014; PMID:1587271
A;Accession: I48299
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 52-247,'L',249-285,'A',287-305,'P',307-416,'S',418-499,'L',501-566,'C',568,'
A;Cross-references: EMBL:X65121; NID:950482; PIDN:CAA46237.1; PID:g667031
A;Reference number: EMBL:X65121; NID:950482; PIDN:CAA46237.1; PID:g667031
J. Biol. Chem. 263, 581-587, 1988
J; Summers, T.A.; Irwin, M.H.; Mayne, R.; Balian, G.
A;Title: Monoclonal antibodies to type X collagen. Biosynthetic studies using an antibod
A;Reference number: S26397; MUID:88087150; PMID:2826450
A;Accession: S26397
A;Molecule type: protein
A;Residues: 'SDGYFSQ',24-26,'KQ' <SUM>
C;Genetics:
A;Gene: Col10a-1
A;Map position: 10
A;Introns: 51/3
C;Superfamily: coiled coil; extracellular matrix; glycoprotein; homotrimer
C;Keywords: signal sequence #status predicted <SIG>
F;1-18/Domain: signal sequence #status predicted <SIG>
F;19-680/Product: collagen alpha 1(X) chain #status predicted <MAT>
F;553-679/Domain: complement C1q carboxyl-terminal homology <C1Q>

Qy   Query Match   6.5%; Score 159; DB 2; Length 680;
     Best Local Similarity 25.5%; Pred. No. 0.0011;
     Matches 97; Conservative 30; Mismatches 108; Indels 146; Gaps 24;

Qy   1 MGPHSGPGGV-RVGALLLLGVLGLVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLD 59
           : :. : ::::. :   :.   :: :   ::
Db   211 IGPP--GFSGVGVRGRENGPFGPGPGI------------------KGDRGFPGEMG------ 244

Qy   60 LLCPRARPPGPHSSPNYEFYKLYLVGGAOGRRCEAPAPNLILTCDRPDLDRFTIKFQE 119
                 :: ::.   :   : :
Db   245 ----PSGPPGPGPP------------GKQGR--EGIGKPGAIGSPGPGI--------- 277

Qy   120 YSPNLWGHBFRSHHDYYIIATSDGT------REGLESLQGGVCLTRGMKVLLRVGQSPRGGA 175
              : : . ::   :: :   :: : ::. . :::
Db   278 --FGEKGHPGSPG-------IAGPPGAPGFPGKQGLPGLRG-----QRG----------PAG-- 314

Qy   176 VPRKPVSEMPMERDRGAAHSLEPGKENLPDGPTSNATSRGARGPLPPPPSMPAVAGAAGGL 235
              .:   :. : .::: : :   : :   : : ::
Db   315 LPGAPGA----KGERGPAG--HPGEPGLPGSP----GNMGPGQGPKGIPGNGIPGAKGEI 364

Qy   236 ALILLGVAGAGGAMCWRRRAKP---SESRHPG------------PGSFGRGGSLGLGGGGG 282
              :: ::. : :   : :::   : ::::::. : : :
Db   365 G--LVGPAGPPGA---RGARGPPGLDGKTGYPGEPGPGLNGPKGNPGLPGQKGDPGVGGTPG 419

Qy   283 M-----------------GRREAEPGELGIALRGGGAAADPPFCHYEKVSGDYGHPVV 323
              :   : : :   : : : :   :   : :
Db   420 LRGPGPVGVGAKGVPGPHGNGBAGPR-GEPGIPGTR-----GPTGPPGVGPGPGSKGDPGNP-- 472

Qy   324 IVQDGP-----------PQSPP 334
              : :   ::
Db   473 -GAPGPAGTATKGLNGPTGPP 492

RESULT 15
CGB07S
collagen alpha 1(III) chain - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 04-Dec-1986 #sequence revision 04-Dec-1986 #text_change 07-May-1999
C;Accession: A02862; A38001; A38002; A38003; A38004; A38005; S71946
R;Fietzek, P.P.; Allmann, H.; Rauterberg, J.; Henkel, W.; Wachter, E.; Kuehn, K.
Hoppe-Seyler's Z. Physiol. Chem. 360, 809-820, 1979
A;Title: The covalent structure of calf skin type III collagen. I. The amino acid sequen
A;Reference number: A02862; MUID:80026026; PMID:488906
A;Accession: A02862
A;Molecule type: protein
A;Residues: 1-242 <FIE>
R;Dewes, H.; Fietzek, P.P.; Kuehn, K.
Hoppe-Seyler's Z. Physiol. Chem. 360, 821-832, 1979
A;Title: The covalent structure of calf skin type III collagen. II. The amino acid sequ

---

A;Reference number: A38001; MUID:80026027; PMID:488907
A;Accession: A38001
A;Molecule type: protein
A;Residues: 243-422 <DEW1>
R;Bentz, H.; Fietzek, P.P.; Kuehn, K.
Hoppe-Seyler's Z. Physiol. Chem. 360, 833-840, 1979
A;Title: The covalent structure of calf skin type III collagen. III. The amino acid seq
A;Reference number: A38002; MUID:80026028; PMID:488908
A;Accession: A38002
A;Molecule type: protein
A;Residues: 423-571 <BEN>
R;Lang, H.; Glanville, R.W.; Fietzek, P.P.; Kuehn, K.
Hoppe-Seyler's Z. Physiol. Chem. 360, 841-850, 1979
A;Title: The covalent structure of calf skin type III collagen. IV. The amino acid sequ
A;Reference number: A38003; MUID:80026029; PMID:488909
A;Accession: A38003
A;Molecule type: protein
A;Residues: 572-808 <LAN>
R;Dewes, H.; Fietzek, P.P.; Kuehn, K.
Hoppe-Seyler's Z. Physiol. Chem. 360, 851-860, 1979
A;Title: The covalent structure of calf skin type III collagen. V. The amino acid seque
A;Reference number: A38004; MUID:80026030; PMID:488910
A;Accession: A38004
A;Experimental source: skin
R;Henkel, W.
Biochem. J. 318, 497-503, 1996
A;Title: Cross-link analysis of the C-telopeptide domain from type III collagen.
A;Reference number: S71946; MUID:96404897; PMID:8809038
A;Accession: S71946
A;Molecule type: protein
A;Residues: 87;106;1017-1029;1037-1049 <HEN>
C;Comment: Prolines at the third position of the tripeptide repeating unit (G-X-Y) are
C;Comment: The type III collagen molecule is a trimer of identical chains, linked to ea
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology
C;Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxypro
F;1-1049/Product: collagen alpha 1(III) chain #status experimental <CAB>
F;1-14/Region: amino-terminal nonhelical telopeptide
F;15-1040/Region: helical
F;587-589/Region: cell attachment (R-G-D) motif
F;752-754/Region: cell attachment (R-G-D) motif
F;875-877/Region: cell attachment (R-G-D) motif
F;878-880/Region: cell attachment (R-G-D) motif
F;935-937/Region: cell attachment (R-G-D) motif
F;1041-1049/Region: carboxyl-terminal nonhelical telopeptide
F;95,107,119,938,950/Modified site: allysine (Lys) #status predicted
F;107,950/Modified site: 5-hydroxylysine (Lys) #status predicted
F;107,950/Binding site: carbohydrate (Lys) (covalent) #status predicted
F;1040,1041/Disulfide bonds: interchain #status experimental

Qy   Query Match   6.3%; Score 154.5; DB 1; Length 1049;
     Best Local Similarity 26.5%; Pred. No. 0.0036;
     Matches 60; Conservative 9; Mismatches 74; Indels 83; Gaps 8;

Qy   171 PRGGAVPRKPVSEMPMERDRGA------------------AHSLEPGKEN 202
            : : : :   :   : :::
Db   688 PAGGSGPAGPPGPGPQGVKGERGSPGGPGAAGPPGGRGPGPPGSNGPGPGSSGPAGKDG 747

Qy   203 LPGDPPSNAT----------SRGAEGPLPPPSMPAVAGAAGGLALILLGVAGA 245
                  : :   : :: :::   : :   :::
Db   748 PPGPPGSNGAPGSPGISGPKGDSGPPGERGAPGPQGGPPGAPGPLGPIG----LTGARGL 802

Qy   246 GGAMCWRRRRAKP-------SESRHPGP------GSFGRGGSLGLGGGGGGMGPRREAEPGE 292
           :: :   : :::   : : : : : :
Db   803 AGPPGMPGARGSPGPGQGIKGBNGKPGPSGQNGERGPPGQGLPGLAGTAGEPGRDGNPGS 862

Qy   293 LGIALRGG--------GAADPPFCPHYEKVSGDYGHPVVYIVQDGPP 330
                |:: ||| |||
Db   863 DGLPGRDGAPGAKGDRGENGSPGAP-------GAPGHP------GPP 896

RESULT 16

CGHU3B

collagen alpha 3(IV) chain precursor, long splice form - human
A;Alternate names: Goodpasture antigen; procollagen alpha 3(IV) chain long splice form
N;Alternate names: Goodpasture antigen; procollagen alpha 1(IV) chain
C;Species: Homo sapiens (man)
C;Date: 28-Oct-1994 #sequence revision 03-Oct-1995 #text_change 22-Jun-1999
C;Accession: A54763; A43928; A44043; A45971; A39786
R;Mariyama, M.; Leinonen, A.; Mochizuki, T.; Tryggvason, K.; Reeders, S.T.
J. Biol. Chem. 269, 23013-23017, 1994
A;Title: Complete primary structure of the human alpha3(IV) collagen chain. Coexpression
R;Reference number: A54763; MUID:94464994; PMID:8083201
A;Accession: A54763
A;Molecule type: mRNA
A;Residues: 1-1670 <MAR>
A;Cross-references: GB:X80031; NID:g577563; PID:g577564
A;Experimental source: kidney
R;Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.
J. Biol. Chem. 267, 19780-19784, 1992
A;Title: Exon/intron structure of the human alpha 3(IV) gene encompassing the Goodpasture
ction.
A;Reference number: A44043; MUID:93015826; PMID:1400291
A;Accession: A44043
A;Molecule type: DNA; mRNA
A;Cross-references: GB:M81379
A;Residues: 1331-1524, /I/,1526-1670 <TUR>
A;Experimental source: kidney
R;Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.
J. Biol. Chem. 269, 17358, 1994
A;Title: Molecular cloning of the human Goodpasture antigen demonstrates it to be the al
A;Reference number: A43928; MUID:9214787; PMID:1737849
A;Accession: A43928
A;Molecule type: mRNA
A;Residues: 1331-1524, /I/,1526-1670 <QUI>
A;Cross-references: GB:M92993; NID:g177895; PIDN:AAA21610.1; PID:g177896
A;Note: sequence extracted from NCBI backbone (NCBIP:1155597)
R;Turner, N.; Mason, P.J.; Brown, R.; Fox, M.; Povey, S.; Rees, A.; Pusey, C.D.
J. Clin. Invest. 89, 592-601, 1992
A;Reference number: A44738; MUID:94274734; PMID:8006044
A;Contents: annotation; erratum; correction to intronic sequence in A44043
R;Bernal, D.; Quinones, S.; Saus, J.
J. Biol. Chem. 268, 12090-12094, 1993
A;Title: The human mRNA encoding the Goodpasture antigen is alternatively spliced.
R;Reference number: A45971; MUID:93280104; PMID:8505332
A;Accession: A45971
A;Molecule type: mRNA
A;Residues: 1427-1444 <BER>
A;Note: sequence extracted from NCBI backbone (NCBIP:133363); sequence incorrectly ident
R;Morrison, K.E.; Mariyama, M.; Yang-Peng, T.L.; Reeders, S.T.
Am. J. Hum. Genet. 49, 545-554, 1991
A;Title: Sequence and localization of a partial cDNA encoding the human alpha3 chain of
A;Reference number: A39786; MUID:91353570; PMID:1882840
A;Accession: A39786
A;Molecule type: mRNA
A;Residues: 1453-1593, /Av, 1595-1670 <MOR>
A;Cross-references: GB:S55790; NID:g234418; PIDN:AAB16637.1; PID:g234419
A;Comment: Prolines and lysines at the third position of the tripeptide repeating unit (
ed and subsequently O-glycosylated.
C;Comment: In Goodpasture's syndrome, an autoimmune response develops against an epitope
A;Genetics:
A;Gene: GDB:COL4A3
A;Cross-references: GDB:128351; OMIM:120070
A;Map position: 2q36-2q37
A;Introns: 1385/1; 1418/1; 1488/1; 1547/2; 1585/3; 1643/2 #status incomplete
A;Note: the alpha 3(IV) and alpha 4(IV) chain genes are encoded on opposite strands with
ed and the alpha 3(IV) type IV collagen is thought to form a heterotrimer of two alpha 3 (
C;Complex: This minor type IV collagen chain is thought to form a heterotrimer of two alpha 3(
mong trimer amino-terminal domains (with disulfide and desmosine cross-links), dimeric a

C;Function:
er associations in the interrupted helical domain (with disulfide and desmosine cross-1
A;Description: minor structural component of extracellular basement membrane in kidney;
A;Superfamily: collagen alpha 1(IV) chain
C;Keywords: alternative splicing; basement membrane; cell binding; coiled coil; extrace
F;1-28/Domain: signal sequence #status predicted <SIG>
F;29-1670/Product: collagen alpha 3(IV) chain, long splice form #status predicted <MAT>
F;29-42/Domain: amino-terminal nonhelical, NH1 <NH1>
F;791-799/Region: interrupted helical
F;791-1438/Region: interrupted helical
F;996-998/Region: cell attachment (R-G-D) motif
F;1154-1156/Region: cell attachment (R-G-D) motif
F;1306-1308/Region: cell attachment (R-G-D) motif
F;1345-1347/Region: cell attachment (R-G-D) motif
F;1432-1434/Region: cell attachment (R-G-D) motif
F;1439-1670/Domain: carboxyl-terminal nonhelical, NC1 <NC1>
F;1491-1551/Domain: collagen IV carboxyl-terminal repeat <CT1>
F;1561-1665/Domain: collagen IV carboxyl-terminal repeat <CT2>
F;31,33,39,41,125,422,476,479,682,722,809,1387/Disulfide bonds: interchain #status pred
F;255/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;1460-1548,1493-1551/Disulfide bonds: (or 1460-1551, 1493-1548) #status predicted
F;1505-1551,1616-1622/Disulfide bonds: #status predicted
F;1570-1662,1604-1665/Disulfide bonds: (or 1570-1665, 1604-1662) #status predicted

Query Match   6.3%; Score 153.5; DB 1; Length 1670;
Best Local Similarity 27.0%; Pred. No. 0.0071;
Matches 108; Conservative 26; Mismatches 155; Indels 111; Gaps 25;

Qy   2 GPPHSGPGG-----VRVGALLLLGVLGLVSGLLSLEPYWNSANKR-----FQAEGGYVLYP   52
             : || || | ||| | || :|| | | :
Db   991 GPP--GPRDGLGSTGNPGEPGLLGIPGSMGNMGMP----GSKGKRGTLGFPGRAGRPGLP 1044

Qy   53 QI-----GDRLDLLCP------RARPPGHSSPYEFYXLYLVGGAQGRRCE---------   93
                     |: || | : || | : | |||
Db 1045 GIHGLQGDKGE---PGYSEGTRPGPGPGTPGDPG-------LPGDMGKKGEMGQPGPGH 1093

Qy   94 ----------APPAPNL--LLTCDRPDLDLRFT-IKFQEYSPNIWGHE----FRSHHDYYII 138
                       |: | | | || | : | |::| :| ||
Db 1094 LGPAGPBGAPGPGSPGLPGKPGPHGDLGFKGILGPPGIRGPPGPGLPGFPGSPGPMGI 1153

Qy   139 ATSDGTRGLESLQGGVCLLTRGMKVLLRVGQSPRG--GAVPRKPVSEMPERDRGAAHSL 196
             : | | | | | | | || : |
Db 1154 RGDGQ--RDGIPGPAG-----EKGETGLLRAPPGPGRGNPGAGQGAK---------GDRGA---- 1196

Qy   197 EPGKKENLPGDPTSNATSRGAEGPLPPSMPAVAGAAGGLALLLLGVAGAGGAMCWRRRRA 256
             :| |: | ||: |||| || | | ::
Db 1197 -PGFPGLPG----RKGAMGDAGPRGPTGIEGFPGFPGPGPLPGA.IIPGQTGNRGP---PGSRG 1248

Qy   257 KPSESRHPGP------------GSFGRGGSLG-LGGGGGMGP--REABPGELGIALRGG 301
                    | || | | | | | | | | | |
Db 1249 SPGAPGPGPGPGSHVIGIKGDKGSMGHPGGMPGPGPRLGAPGTPGLP---GP 1305

Qy   302 AAADPPFCPHYEKVSGDYGHPVVYIVQDGP-----PQSPPNI 336
             ||||||||||||||||||||||||||||| |:|| :
Db 1306 RGDGP--QGFPGVKGEKGNPGFLGSIGPPGPGPIGPKGPPGV 1344

RESULT 17

B31219

collagen 2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 30-Jun-1989 #sequence revision 30-Jun-1989 #text_change 21-Jul-2000
C;Accession: B31219; T37289; T26033; T37288
R;Kramer, J.M.; Cox, G.N.; Hirsh, D.
Cell 30, 599-606, 1982
A;Title: Comparisons of the complete sequences of two collagen genes from Caenorhabditi
R;Reference number: A90826; MUID:83050944; PMID:7139711
A;Accession: B31219
A;Molecule type: DNA
A;Residues: 1-301 <KRA>
A;Cross-references: GB:V00148; NID:g6683; PIDN:CAA23464.1; PID:g6684
A;Accession: T37289
A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA
A;Residues: 1-301 <KR3>
A;Cross-references: EMBL:V00148; NID:g6683; PIDN:CAA23464.1; PID:g6684
R;Coles, L.
submitted to the EMBL Data Library, December 1995
A;Reference number: Z20141
A;Accession: T26033
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-301 <WIL>
A;Cross-references: EMBL:Z68301; PIDN:CAA92620.1; GSPDB:GN00022; CESP:W01B6.7
A;Experimental source: clone W01B6
A;Residues: 1-301 <WIL>
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-301 <KR2>
R;Kramer, J.M.; Cox, G.N.; Hirsh, D.
J. Biol. Chem. 260, 1945-1951, 1985
A;Title: Expression of the Caenorhabditis elegans collagen genes col-1 and col-2 is deve
A;Reference number: Z21668; MUID:85105075; PMID:2578467
A;Accession: T37288
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-301 <KR2>
A;Cross-references: EMBL:J01048; NID:g156261; PIDN:AAA27990.1; PID:g156262
C;Genetics:
A;Gene: col-2
A;Map position: 4
A;Introns: 94/1
C;Superfamily: unassigned collagens

Query Match   6.2%;   Score 151.5; DB 2;   Length 301;
Best Local Similarity   28.7%;   Pred. No. 0.0015;
Matches   50;   Conservative   15;   Mismatches   72;   Indels   37;   Gaps   7;

```
Qy  168 GQSPRGGAVPRKPVSEMPMERDRG------------AAHSLEPGKENLPGDPT--SNATSRGAE 217
Db  129 GNPGKGASAPCEPVTQPPCQPCGPGPGPAGPPGPPGPGDGNPGGSPAGPSGPGPAGPP 188

Qy  218 GPFLPPPSMPAVAGAAGGLALLLGVAGA----GGAMCWRRRAKPSESRHPGP--GSFGRGG 273
Db  189 GPAGPAGNDGAPGAPGG----PGEPGASEQGGP----GEPGPAGPPGPAGPAGNDG 236

Qy  274 SLG|LGGGGGMGPR-------------EAEPGELGIALRGGGAADPPFCHYEKYSG 316
Db  237 APGTGGPGPAGPKGPPGPAGAGPAGADGNPGGPGTPAGKPGGEGEKGICPKYCAIDG 290
```

RESULT 18
T32248
hypothetical protein T15B7.4 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 21-Jan-2000
A;Accession: T32248
R;Pauley, A.; Gattung, S.
submitted to the EMBL Data Library, September 1997
A;Description: The sequence of C. elegans cosmid T15B7.
A;Reference number: Z21139
A;Accession: T32248
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-325 <PAU>
A;Cross-references: EMBL:AF022985; PIDN:AAB69961.1; GSPDB:GN00023; CESP:T15B7.4
A;Experimental source: strain Bristol N2; clone T15B7
A;Accession: T32248
A;Gene: CESP:T15B7.4
A;Map position: 5
A;Introns: 266/1
C;Superfamily: unassigned collagens

Query Match   6.1%;   Score 149; DB 2;   Length 325;
Best Local Similarity   26.0%;   Pred. No. 0.0024;
Matches   92;   Conservative   25;   Mismatches 109;   Indels 128;   Gaps   23;

```
Qy   13 VGALILLGVLGLVSGLSL--------EPVWNSANKRFQAEGG-----YVLYPQIGDRLDLIC  62
Db    1 MSASTLVTVASASGIAIVVCVFTVGMIFNDINSFYDEKIGELKEFGYEQIA--WQAMI  58
```

Qy   63 PRARPPGPHSSPNYEFYKLYLVG--------------GAQGRRCEA-PPAPNLLLTCDRPDLD; 111
Db   59 PTTRPSSGSS--------------FLLGRNKRQAQCNCGAQSRGCPAGPGP--------- 96

Qy 112 RFTIKFQEYSPNLWGHEFRSHHDYIIATSDGTRGLESILGQ--------- 169
Db   97 ----PCQPGAPGGBQGHPG------LAQGPGSGARINPATGRPGFCIT------------ 133

Qy 170 SPRGGAVPRKPVSEMPMERDRGAAHSLEPGKENLPGDPTSNATSRGAEGPLPPPSMPAVA 229
Db 134 CPPAGAGPPAGP--------PGFKGNNGQPGAPAQS-GGRGP-PGPRGP--A 175

Qy 230 GAAGGLALLLGVAGAGGAMCWRRRAKPSESRHPG-PGSFGRGG--SLGLGGGGGM--- 283
Db 176 GDAG-----------------SPGGQPGHPGSPGNPGRGGGRSRGLPGPSGSGRPGP 212

Qy 284 -GPREAPPGELGIALRGGGAADP-PFCPHYEKVSGDYGHPVYIVQDGPPQSPPN 335
Db 213 PGP-AGGPGQPG---HSGGASGSPGPGGP-----PGPSGQPGHSGNDGVPGAPGN 257

RESULT 19
S42886
collagen - silkworm
C;Species: Bombyx mori (silkworm)
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 20-Sep-1999
A;Accession: S42886
R;Chareyre, P.P.; Besson, M.M.; Fourche, J.J.; Bosquet, G.G.
submitted to the EMBL Data Library, March 1994
A;Description: A novel collagen encoding mRNA is transiently accumulated during ecdysor
A;Reference number: S42886
A;Accession: S42886
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-569 <CHA>
A;Cross-references: EMBL:Z30348; NID:g457768; PIDN:CAA83002.1; PID:g457769
C;Superfamily: unassigned collagens

Query Match   6.1%;   Score 149; DB 2;   Length 569;
Best Local Similarity   29.3%;   Pred. No. 0.0045;
Matches   51;   Conservative   14;   Mismatches   77;   Indels   32;   Gaps   6;

Qy 171 PRGGAVPRKPVSEMPMERDRGAAHSLEPGKENLPGDPTSNATSRGAABGPLPPPSMPAVAG 230
Db   1 PGQGQGQPIKGQ--------PGYPGQPGQPGYPGQP-CQPGAPGAPGQPGQPGQPGTPG 50

Qy 231 AAGGLALLLGVAGAGGAMCWRRRAKPSESRHPG--------PGSFGRGGSLGLGGGGGM 283
Db   51 QAG-----QPGYPGQG-----QPIKPAQPGHPGQPGQPGQPGQPGQPGYPGQPGPQ 99

Qy 284 GPREAPPGELGIALRG-----GGAADPPFCHYEKVSGDYGHPVYIVQDGPPQSP 333
Db 100 PGQPGQPGQGYPGQGGPGQGPINPSGPGYPGQGGPGQPGYPGQPVQPGQPGQP 153

RESULT 20
T20497
hypothetical protein F02D10.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
A;Accession: T20497
R;Swinburne, J
submitted to the EMBL Data Library, November 1995
A;Reference number: Z19283
A;Accession: T20497
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-316 <WIL>
A;Cross-references: EMBL:Z67990; PIDN:CAA91932.1; GSPDB:GN00028; CESP:F02D10.1
A;Experimental source: clone F02D10
C;Genetics:
A;Gene: CESP:F02D10.1

A;Map position: X
A;Introns: 56/3
C;Superfamily: unassigned collagens

Query Match
Best Local Similarity 28.5%; Pred. No. 0.0025; DB 2; Length 316;
Matches 49; Conservative 12; Mismatches 78; Indels 33; Gaps 6;

Qy 168 GQSPRGGAVPRKPVSEMPMERDRGAAHSLEPGKENLPGDP-------------------TSNATSRGA 216
Db 144 GASGKGAAAPCEPSTPPPCCQ-PCPAGPGPGPGPDGTPGEPGGPGPAGSPAGPGPGPAGP 202

Qy 217 EGPLPPPSMPAVAGAAGGLALLLLGVAGAGGAMCWRRRAKPSESRHPGP-GSFGRGGSL 275
Db 203 PGPAGPPGNDGQPGPGPG--QDGASSAGG------EAGPPGAGPPGPAGPDGQS 253

Qy 276 GLGGGGGMGPR------------EAEPGELGIALRGGGAADPPFCPHYEKVSG 316
Db 254 GSGSAGGPGPGPKGPGPAGQPGSDGNPGTAGPPGNPGGEGEBGKGICPKYCAIDG 305

RESULT 21
S42617
collagen alpha 1(IX) chain - mouse
collagen alpha 1(IX) chain (house mouse)
C;Species: Mus musculus (house mouse)
C;Date: 25-Dec-1994 #sequence_revision 19-Apr-1996 #text_change 20-Sep-1999
C;Accession: S42617
R;Rokos, I.; Muragaki, Y.; Warman, M.; Olsen, B.R.
Matrix Biol. 14, 1-8, 1994
A;Title: Assembly and sequencing of a cDNA covering the entire mouse alpha-1(IX) collage
A;Reference number: S42617; MUID:94340199; PMID:8061915
A;Accession: S42617
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-921 <ROK>
A;Cross-references: GB:L22215; GB:L19396; NID:g192675; PIDN:AAA21834.1; PID:g192676
C;Superfamily: unassigned collagens

Query Match
Best Local Similarity 24.0%; Pred. No. 0.009; DB 2; Length 921;
Matches 87; Conservative 23; Mismatches 130; Indels 122; Gaps 17;

Qy 2 GPPHSGPGPGGVRVGALLLLGVLGLVSGLSLEPVIWNSANKRFQAEGGYVLYPQIGDRLDLL 61
Db 391 GPP-------GPPGPS-------GTIGFHDG 410

Qy 62 CPRARPPGFHSSPNYEFYKLYLVG-------GAQGRRCEAPPAPNLLLTCDRPDLLRFTIK 116
Db 411 CPNSCPPGRSGYPG-------LPGMRGHKGAKGEIGE 440

Qy 117 FQEYSPNLWGHEFRSHHDYYIIATSDGTREGLESIQG---GVCLTRGMKVLLRVSQSPRG 174
Db 441 -----PGRQCHKGEEGDQGELGEVGDQGPGPGQLRGRGITGIVGDKGEKGARGFDGEPGPGPQ 495

Qy 175 AVP---------RKPVSEMPMERDRGAAHSLEPGKENLPGDPTSNATS--RGAEGPLPPPS 224
Db 496 GIPGAGDQGSCRGPPGEETGFPGEDRGI-------QSGRGIPGSPGPRKGDTGLPGVDGRDGIPG 550

Qy 225 MPAVAGAAG----------GLALL--LLGVAGAGGAMCWRRRARKPSE-------SRHPG--- 265
Db 551 MPGTKGEBAGKPGPGPGDVGIAGLPGVPGIPGAKGVAGEKGNTGAPGKPGPCQLGSSGKKPGQQ 610

Qy 266 -PGSFGRGGSLGL-GGGGMGPREAB----PGELGIALRGG--GAADPPFCPHYEKVSGDYG 319
Db 611 PPGEVGPRGRGLIPGSRGPVGPEGSPGIPGKLGSVGSPGLPGLPGPPGLPGMKGDRGVYG 670

Qy 320 HP 321
Db 671 BP 672

RESULT 22
A56101

collagen alpha 1 (XVIII) chain precursor, short splice form - mouse
N;Contains: endostatin
C;Species: Mus musculus (house mouse)
C;Date: 03-Oct-1995 #sequence_revision 08-May-1998 #text_change 31-Mar-2000
C;Accession: A56101; A58371; S72450; S65595
R;Rehn, M.; Pihlajaniemi, T.
J. Biol. Chem. 270, 4705-4711, 1995
A;Title: Identification of three N-terminal ends of type XVIII collagen chains and tissu
tif homologous to rat and Drosophila frizzled proteins.
A;Reference number: A56101; MUID:95181468; PMID:7876242
A;Accession: A56101
A;Molecule type: mRNA
A;Residues: 1-103 <REH1>
A;Cross-references: GB:U11636; NID:g618427; PIDN:AAC52178.1; PID:g618428
R;Rehn, M.; Pihlajaniemi, T.
Proc. Natl. Acad. Sci. U.S.A. 91, 4234-4238, 1994
A;Title: Alpha1(XVIII), a collagen chain with frequent interruptions in the collagenous
A;Reference number: A58371; MUID:94240112; PMID:8183894
A;Accession: A58371
A;Molecule type: mRNA
A;Residues: 1-928 <REH2>
A;Cross-references: GB:L16698; NID:g404754; PIDN:AAA37434.1; PID:g553894
R;Oh, S.P.; Kamagata, Y.; Timmons, S.; Ooshima, A.; Olsen, B.R.
submitted to the EMBL Data Library, August 1993
A;Reference number: S72450
A;Accession: S72450
A;Molecule type: mRNA
A;Residues: 28-687,'L',689-734,'F',736-751,'R',753-1315 <OHW>
A;Cross-references: EMBL:L25545; NID:g348968; PIDN:AAA19787.1; PID:g511298
R;Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, A.B.R.
Proc. Natl. Acad. Sci. U.S.A. 91, 4229-4233, 1994
A;Title: Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-Xaa
A;Reference number: A58370; MUID:94240111; PMID:8183893
A;Accession: S65595
A;Molecule type: mRNA
A;Residues: 28-1315 <OHS>
A;Cross-references: EMBL:L22545
A;Comment: Prolines and lysines at the third position of the tripeptide repeating unit
lated and subsequently O-glycosylated.
C;Comment: The different splice forms of collagen alpha 1(XVIII) may be involved in peri
C;Comment: Endostatin is released from collagen alpha 1(XVIII) chain by the action of ur
ay be useful in treating solid tumors.
C;Genetics:
A;Gene: MGI:Col18a1
A;Cross-references: MGI:71175
A;Map position: 10:41.0
C;Superfamily: unassigned collagens
C;Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteoglyc
F;1-25/Domain: signal sequence #status predicted <SIG>
F;24-235/Region: thrombospondin amino-terminal similarity
F;26-1315/Product: collagen alpha 1(XVIII) chain, short splice form #status predicted <N
F;327-353/Domain: collagenous #status predicted <C01>
F;364-437/Domain: collagenous #status predicted <C02>
F;462-583/Domain: collagenous #status predicted <C03>
F;607-689/Domain: collagenous #status predicted <C04>
F;704-745/Domain: collagenous #status predicted <C05>
F;759-831/Domain: collagenous #status predicted <C06>
F;842-874/Domain: collagenous #status predicted <C07>
F;887-910/Domain: collagenous #status predicted <C08>
F;892-894/Region: cell attachment (R-G-D) motif
F;918-969/Domain: collagenous #status predicted <C09>
F;983-1000/Domain: collagenous #status predicted <C10>
F;1132-1315/Product: endostatin #status predicted <EST>
F;1139-1315/Region: multiplexin collagen carboxyl-terminal similarity
F;126,488/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;172-226/Disulfide bonds: carbohydrate #status predicted
F;240,245,1257/Binding site: carbohydrate (Ser) (covalent) #status predicted
F;451,454,594/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match
Best Local Similarity 23.7%; Pred. No. 0.017; DB 2; Length 1315;
Matches 92; Conservative 22; Mismatches 107; Indels 167; Gaps 21;

```
Qy      2 GPPHSGPGGVRVGALLLGVLG-LVSGLSLEPYVWNSANKRFQAEGGYVLYPQIGDRLDL 60
            ::                 ::           :    :    :
Db    339 GPP--GPAGPQ-----------GPAGPVVQSPNSQPV-------------------------- 362

Qy     61 LCPRAR-PPGPHSSPNYEFYKLYLVGAQGRRCE--------------------------- 93
                ::                  :       :
Db    363 --PGAQGPPGPQGGPPGKD--------GTPGRDGEPGDPDGEDGRPGDTGPQGFPGTPGDVG 412

Qy     94 -----------APPAPNLLLTCDRPDLDLRFTIKFQEYSPNLWGHEFRSHDHYYIIAT 140
                                  :                 :
Db    413 PKGEKGDPGIGPRGPPGP-----------PGPGPGPSFRQDKLTFIDME-GAAHSLE 449

Qy    141 SDGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKP--VSEMPMERDR-GAAHSLE 197
                  :            :              :
Db    450 GSGPFSGGDIESLRG-----------PRGPPGPGPGPGVPGLPGEPGPRGPGINGSYA 492

Qy    198 PGKENLPGDPTSNATSRGAEGP--LP-PPSMPAVAGAAGGLALLLLGVAGGAGMCWRRR 254
              :            :                 :        :
Db    493 PGPAGLPGVP----------GKEGPGPGPGPPGPPGGPPGPGGPGPGKEGP------PGVAGQGKGSV 535

Qy    255 RAKPSESRHPGP-GSFGRGGSLGLGGGGMGPREAEPGELGIALRGG--GAADPPFCPHY 311
             :            :                :
Db    536 -------GDVGIPGPKGSKGDLGP2IGMPGKSGLA---GSFGPVGPPGPPGPGPGPGPFAAGF 588

Qy    312 EKVSGDYGHPVVIVQ------DGPPQSP 333
              :              :
Db    589 DDMEGS-GIPLWTTARSSDGLQQPPGSP 615
```

RESULT 23

```
A40333
collagen alpha 1(II) chain precursor - African clawed frog
C;Species: Xenopus laevis (African clawed frog)
C;Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 16-Jul-1999
C;Accession: A40333
R;Su, M.W.; Suzuki, H.R.; Bieker, J.J.; Solursh, M.; Ramirez, F.
J. Cell Biol. 115, 565-575, 1991
A;Title: Expression of two nonallelic type II procollagen genes during Xenopus laevis em
A;Reference number: A40333; MUID:92011898; PMID:1918153
A;Molecule type: mRNA
A;Status: nucleic acid sequence not shown
A;Residues: 1-1492 <SUA>
A;Molecule type: mRNA
Matches 56; Conservative 11; Mismatches 80; Indels 33; Gaps 8;
Best Local Similarity    31.1%; Pred. No. 0.02;
Query Match    6.0%; Score 146.5; DB 2; Length 1492;

A;Note: this sequence is presented as substitutions relative to another sequence in a fi
es they replace; the appropriate interpretation of the sequence was reconstructed
A;Cross-references: GB:M63596
C;Superfamily: collagen alpha 1(t) chain; fibrillar collagen carboxyl-terminal homology;
C;Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix
F;37-96/Domain: von Willebrand factor type C repeat homology <VWC>
F;1263-1492/Domain: fibrillar collagen carboxyl-terminal homology <FCC>
```

```
Qy    171 PRGGAVPRKPVSEMPMERDRG--AAHSLEPGKENLPGDPTSNATSRGAEGPLPPPSMPAVA 229
                                       :              :
Db    123 PRGPPGPGPGPGPSGEQGPGERGDKGETGAPGPGRCRDGEP---GTPGNPGPGPGPPGPP 178

Qy    230 GAAGGLALLLLG--VAGAGGAMCWRRR----------RAKPSESRHPGPGSF-GRGGSLGLG 278
             :                                  :
Db    179 GLGGNFAAQMTGFDEKAGGAQWGVAMQGPMGPMGPRGPRGPGPGSGAPGPGPQGFRQGNPGEPGES 238

Qy    279 GGGG-MGPR-----------EAAFGELGIALRGG----------GAADPPFCPHYEKVSGDYGHP 321
             :                                  :
Db    239 GAGGPMGPGPRGPGPGPSGKPGDDGBAGKPGKSGERGPGPQGPGARGFPGTPGPLPGVKGHRGYP 298
```

RESULT 24

```
B56101
collagen alpha 1(XVIII) chain precursor, long splice form - mouse
collagen alpha 1(XVIII) chain precursor, medium splice form; endostatin
N;Contains: collagen alpha 1(XVIII) chain precursor, medium splice form #
C;Species: Mus musculus (house mouse)
```

(right column, continued annotation blocks)

```
C;Date: 03-Oct-1995 #sequence_revision 08-May-1998 #text_change 31-Mar-2000
C;Accession: B56101, C56101; S72450; S65595; PN0675; A54072; A58816
R;Rehn, M.; Pihlajaniemi, T.
J. Biol. Chem. 270, 4705-4711, 1995
A;Title: Identification of three N-terminal ends of type XVIII collagen chains and tiss
tif homologous to rat and Drosophila frizzled proteins.
A;Reference number: A56101; MUID:95181468; PMID:7876242
A;Molecule type: mRNA
A;Residues: 1-562 <REH1>
A;Cross-references: GB:U11637; NID:g618429; PIDN:AAC52179.1; PID:g618430
A;Experimental source: splice form clone PE17.24
A;Accession: C56101
A;Molecule type: mRNA
A;Residues: 1-239, 487-562 <REH2>
A;Cross-references: GB:U11637; NID:g618429
A;Experimental source: splice form clones PE8.1, PE19, PE15.2
A;Accession: S72450
A;Molecule type: mRNA
R;Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, B.R.
submitted to the EMBL Data Library, August 1993
A;Reference number: S72450
A;Accession: S72450
A;Molecule type: mRNA
A;Residues: 487-1146, 'L', 1148-1193, 'F', 1195-1210, 'R', 1212-1512, 'L', 1514-1522, 'F', 1524-1
R;Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, A.B.R.
Proc. Natl. Acad. Sci. U.S.A. 91, 4229-4233, 1994
A;Title: Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-Xa
A;Reference number: A58370; MUID:94240111; PMID:8183893
A;Accession: S65595
A;Molecule type: mRNA
A;Residues: 487-1512, 'L', 1514-1522, 'F', 1524-1683, 'V', 1685-1774 <OH2>
A;Cross-references: EMBL:L22545
R;Abe, N.; Muragaki, Y.; Yoshioka, H.; Inoue, H.; Ninomiya, Y.
Biochem. Biophys. Res. Commun. 196, 576-582, 1993
A;Title: Identification of a novel collagen chain represented by extensive interruptio
A;Reference number: PN0675; MUID:94059075; PMID:8240330
A;Accession: PN0675
A;Molecule type: mRNA
A;Residues: 635-1774 <ABB>
R;Rehn, M.; Hintikka, E.; Pihlajaniemi, T.
J. Biol. Chem. 269, 13929-13935, 1994
A;Title: Primary structure of the alpha1 chain of mouse type XVIII collagen, partial st
collagen chain.
A;Reference number: A54072; MUID:94245707; PMID:8188673
A;Accession: A54072
A;Molecule type: DNA; mRNA
A;Residues: 1293-1403, 'R', 1405-1774 <REH3>
A;Cross-references: GB:U03714; NID:g487733; PIDN:AAA20657.1; PID:g487734
R;O'Reilly, M.S.; Boehm, T.; Shing, Y.; Fukai, N.; Vasios, G.; Lane, W.S.; Flynn, E.; B
Cell 88, 277-285, 1997
A;Title: Endostatin: an endogenous inhibitor of angiogenesis and tumor growth.
A;Reference number: A58816; MUID:97160848; PMID:9008168
A;Accession: A58816
A;Molecule type: protein
A;Residues: 1551-1610 <ORE>
A;Experimental source: hemangioendothelium cells
A;Note: inhibits endothelial cell proliferation
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit
lated and subsequently O-glycosylated.
C;Comment: The different splice forms of collagen alpha 1(XVIII) may be involved in per
C;Comment: Endostatin is released from collagen alpha 1(XVIII) chain by the action of u
ay be useful in treating solid tumors.
C;Genetics:
A;Gene: MGI:Col18a1
A;Cross-references: MGI:71175
A;Map Position: 10:41.0
A;Introns: 1295/3; 1310/1; 1331/1; 1345/3; 1388/3; 1437/1; 1461/3; 1505/3; 1516/3; 1599
A;Note: the list of introns is incomplete
C;Superfamily: unassigned collagens
C;Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteogly
F;1-1774/Product: collagen alpha 1(XVIII) chain precursor, long splice form #status pre
F;1-239,487-1774/Product: collagen alpha 1(XVIII) chain precursor, medium splice form #
F;1-24/Domain: signal sequence #status predicted <SIG>
```

```
F;361-486/Region: frizzled similarity
F;786-812/Domain: collagenous #status predicted <CO01>
F;823-896/Domain: collagenous #status predicted <CO02>
F;921-1042/Domain: collagenous #status predicted <CO03>
F;1066-1148/Domain: collagenous #status predicted <CO04>
F;1163-1204/Domain: collagenous #status predicted <CO05>
F;1218-1290/Domain: collagenous #status predicted <CO06>
F;1301-1333/Domain: collagenous #status predicted <CO07>
F;1346-1369/Domain: collagenous #status predicted <CO08>
F;1351-1353/Region: cell attachment (R-G-D) motif
F;1377-1428/Domain: collagenous #status predicted <CO09>
F;1442-1459/Domain: collagenous #status predicted <CO10>
F;1591-1774/Product: endostatin #status predicted <EST>
F;1598-1774/Region: multiplexin collagen carboxyl-terminal similarity
F;354,361,947/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;699,704,1716/Binding site: carbohydrate (Ser) (covalent) #status predicted
F;1591-1774/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted
F;910,913,1053/Binding site: ...
```

```
Query Match    6.0%;  Score 146.5;  DB 2;  Length 1774;
Best Local Similarity  23.7%;  Pred. No. 0.024;
Matches  92;  Conservative  22;  Mismatches  107;  Indels  167;  Gaps  21;

Qy    2 GPPHSGPGGVRVGALLLLGVLG-LVSGLSLEPYVWNSANKRFQABGGYVLYPIGDRLDL 60
Db  798 GPP--GPAGPQ----------GPAGPVVQSPNSQPV-------------------- 821

Qy   61 LCPRAR-PGGPHSSPNYEFYKLYLVGGAQGRRCE------------------ 93
Db  822 --PGAAQGPGPQGPPGKD-------GTPGRDGEPGDPGDBDGRPGDTGPGPGPGTPGDVG 871

Qy   94 ------------APPAPNLLLTCDRPDLDLRFTIKFQEYSPNLWGHEPRSHHDYYIIAT 140
Db  872 PKGEKGDPGIGPRGPPGP------------------PGPPGPSPRQDKLTFIDME 908

Qy  141 SDGTREGLSLSLQGGVCLTRGMKVLLRVCQOSPRGGAVPRKP--VSEMPMERDR-GAAHSLE 197
Db  909 GSGFBGDIESLRG---------------PRGPPGPGPPGVPGLPGSPGRFGINGSYA 951

Qy  198 PGKENLPGDPTSNATSRGAEGP--LP-PPSMPAVAGAAGGLALLLIGVAGAGAMCWRRR 254
Db  952 PGPAGLPGVP----------GKEGPGPPGPPGPPGPPGKEGP------PGVAGQKGSV- 994

Qy  255 RAKPSESRHPGP-GSFGRGSLGLGGGGGMGPREAEPGELGIALRGG--GAADPFCPHY 311
Db  995 -----GDVGIPGPKGSKGDLGPIGMPGKSGLA---GSPGPVGPPGGPGPPGPGPGFAAGF 1047

Qy  312 EKVSGDYGHPVYIVQ-----DGPPQSP 333
Db 1048 DDMEGS-GIPLWTTARSSDGLQGPPGSP 1074
```

RESULT 25
S20819
collagen alpha 3(IX) chain precursor - chicken
C;Species: Gallus gallus (chicken)
C;Date: 22-Nov-1993 #sequence_revision 09-Mar-1996 #text_change 20-Sep-1999
C;Accession: S20819; S22429; S22918; S22238; C18856; S22241
R;Brewton, R.G.; Ouspenskaia, M.V.; van der Rest, M.; Mayne, R.
Eur. J. Biochem. 205, 443-449, 1992
A;Title: Cloning of the chicken alpha-3(IX) collagen chain completes the primary structu
A;Reference number: S20819; MUID:92241276; PMID:1572350
A;Molecule type: mRNA
A;Accession: S20819
A;Residues: 1-675 <BRE>
R;Har-El, R.; Sharma, Y.D.; Aguilera, A.; Ueyama, N.; Wu, J.J.; Eyre, D.R.; Juricic, L.;
J. Biol. Chem. 267, 10070-10076, 1992
A;Title: Cloning and developmental expression of the alpha3 chain of chicken type IX col
A;Reference number: S22429; MUID:92250566; PMID:1577778
A;Molecule type: mRNA
A;Residues: 1-195, 'G', 197-675 <HA1>
A;Cross-references: EMBL:M83179

A;Note: 353-Arg, 386-Leu and 548-Arg were also found
R;Har-El, R.; Sharma, Y.D.; Aguilera, A.; Ueyama, N.; Wu, J.J.; Eyre, D.R.; Juricic, L.;
submitted to the EMBL Data Library, February 1992
A;Description: Cloning and developmental expression of the alpha 3 chain of chicken typ
A;Reference number: S22918
A;Accession: S22918
A;Molecule type: mRNA
A;Residues: 1-195, 'G', 197-405, 'S', 407-675 <HA2>
A;Cross-references: EMBL:M83179; NID:g211040; PIDN:AAB59960.1; PID:g211041
R;Mayne, R.; van der Rest, M.; Ninomiya, Y.; Olsen, B.R.
Ann. N. Y. Acad. Sci. 460, 38-46, 1985
A;Title: The structure of type IX collagen.
A;Reference number: S22238; MUID:86185164; PMID:3868958
A;Accession: S22238
A;Molecule type: protein
A;Residues: 135, E',137-187,'X',189-191 <SHI>
A;Superfamily: unassigned collagens
C;Keywords: coiled coil; connective tissue; disulfide bond; extracellular matrix; heter
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-675/Product: collagen alpha 3(IX) chain #status predicted <MAT>
F;22-24/Domain: non-collagenous NC4 #status predicted <NC4>
F;25-161/Domain: collagenous COL3 #status predicted <COL3>
F;162-176/Domain: non-collagenous NC3 #status predicted <NC3>
F;177-515/Domain: collagenous COL2 #status predicted <COL2>
F;516-546/Domain: non-collagenous NC2 #status predicted <NC2>
F;547-657/Domain: collagenous COL1 #status predicted <COL1>
F;658-675/Domain: non-collagenous NC1 #status predicted <NC1>
F;22/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicte
F;137,143,146,149,152,155,179,182,185,552,558/Modified site: hydroxyproline (Pro) #stat
F;170,174,525,658,663/Disulfide bonds: interchain #status predicted
R;Ninomiya, Y.; van der Rest, M.; Mayne, R.; Lozano, G.; Olsen, B.R.
Biochemistry 24, 4223-4229, 1985
A;Title: Construction and characterization of cDNA encoding the alpha2 chain of chicken
A;Reference number: A18856; MUID:86026268; PMID:2996593
A;Accession: C18856
A;Molecule type: protein
A;Residues: 540-558 <NIN>
R;Shimokomaki, M.; Wright, D.W.; Irwin, M.H.; van der Rest, M.; Mayne, R.
Ann. N. Y. Acad. Sci. 580, 1-7, 1990
A;Title: The structure and macromolecular organization of type IX collagen in cartilage
A;Reference number: S22241; MUID:90247791; PMID:2186687
A;Accession: S22241
A;Molecule type: protein
A;Residues: 540-548 <MAY>
```

```
Query Match    6.0%;  Score 146;  DB 2;  Length 675;
Best Local Similarity  28.8%;  Pred. No. 0.0089;
Matches  64;  Conservative  9;  Mismatches  71;  Indels  78;  Gaps  12;

Qy  166 RVGQSPRGGAVPRKPVSEMPMERDRGAAHSLEPGKENLPGDPTSNATSRGAEGPLPPPSM 225
Db   23 RVG--PQGPPGPGPPGPGPSGKDGIDG------EPGPSGLPGPP-------GPKGAPGK 65

Qy  226 PAVAGAAG--GLALLLGVAGAGGAMCWRRRRAKPSESRPG------------PGSPGRG- 272
Db   66 PGAAGEAGLPGLP-----GVDGLTGT------DGPGPNGPGDRGALGPAGPPGPGPAGKGL 115

Qy  273 ------GSLGLGGGGGM---GPR-----------EABGGBLGIALRGG-------- 300
Db  116 PGPPGPPGPGPSGLPGGNGFRGPPGPGPPGPGPPGPQLAGIIPBGGGDLQCPALCP 175

Qy  301 -GAADPFCPHYEKVSGDYGHPVYIVQDGP-----PQSPPNI 336
Db  176 PGPPGPPGPMGFGFKGHTGHKGEPGEGPGEGKEGEKGSPGPGPPGPGI 217
```

RESULT 26
T20906
hypothetical protein F14F7.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C;Accession: T20906
R;McMurray, A.

submitted to the EMBL Data Library, November 1996
A;Reference number: Z19345
A;Accession: T20906
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-305 <WIL>
A;Cross-references: EMBL:Z81503; PIDN:CAB04111.1; GSPDB:GN00021; CBSP:F14F7.1
A;Experimental source: clone F14F7
A;Genetics:
A;Gene: CESP:F14F7.1
A;Map position: 3
A;Introns: 273/3, 49/3
C;Superfamily: unassigned collagens

Query Match
Best Local Similarity 22.5%; Pred. No. 0.004; Score 145.5; DB 2; Length 305;
Matches 83; Conservative 21; Mismatches 122; Indels 143; Gaps 12;

```
Qy   11 VRVGALLILG------VLGLVSGLSLEPVYWN------ 40
         ||:|| :                 |:|
Db    5 VRLKAYRLLAFSAVAFCVLSVSVCITLPMVYNYVSGMRVRLVQEVSFCKHSANEVFTEV 64

Qy   41 ------RPQAEGGYVLYPQIGDRDLLCPRARPPGHSSPNYEFYKLYLVGG 86
Db   65 NHLPASATNSTRSARHAGYGGY--------------AQPQGGGG------GG 97

Qy   87 AQGRRCEAPPAPNLLLTCDRPDLDLRFTIKFQEYSPNLWGHEFRSHHDYYIIATSDGTRE 146
Db   98 GQCSTCRPGPGPGAGGTPGKPG----------------RPGAPGAWG---------- 128

Qy  147 GLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKPVSEMPERDRGAAHSLEPGKENLPGD 206
Db  129 ------MPGNPGKGGSGPCHPVVPTPCKPCPGG----RPGPGPPGR 165

Qy  207 ---ATSRGA------------EGPLPPPSMPAVAGAAGGLALLLLGVAGAGGAMCW 251
Db  166 PGSDGQPGRPATGGGAAPRPGPPGGPXPGPGAPCGNSGRAGAPGQPGNDAHGYGGGVGAPGP 225

Qy  252 RRRAKPSESRHPGPGSPGRGGSLGLGGGGGM------GPREAEPGELGIALRGGGAADPPF 307
Db  226 AGPRGAPGPAGHPGSSGGGRGPGPAGPGPAGPGQPGRPGP-DGHPGQPGRPGQSGSGSNRGV 284

Qy  308 CPHYEKVSG 316
Db  285 CPKYCAIDG 293
```

RESULT 27
S13301
collagen alpha 1(X) chain precursor - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 21-Nov-1993 #sequence_revision 23-Feb-1996 #text_change 13-Aug-1999
C;Accession: S13301
R;Thomas, J.T.; Kwan, A.P.L.; Grant, M.E.; Boot-Handford, R.P.
Biochem. J. 273, 141-148, 1991
A;Title: Isolation of cDNAs encoding the complete sequence of bovine type X collagen. Ev
A;Reference number: S13301; MUID:91113131; PMID:1703407
A;Accession: S13301
A;Molecule type: mRNA
A;Residues: 1-674 <THO>
C;Cross-references: EMBL:X53556; NID:g9263; PIDN:CAA37624.1; PID:g9264
C;Genetics:
C;Gene: COL10A1
C;Superfamily: collagen alpha 1(X); complement C1q carboxyl-terminal homology
C;Keywords: coiled coil; extracellular matrix; glycoprotein; homotrimer
F;1-18/Domain: signal sequence #status predicted <SIG>
F;19-674/Product: collagen alpha 1(X) chain #status predicted <MAT>
F;547-673/Domain: complement C1q carboxyl-terminal homology <C1Q>

Query Match
Best Local Similarity 24.2%; Score 145; DB 2; Length 674;
Matches 100; Conservative 30; Mismatches 125; Indels 158; Gaps 24;

```
Qy   15 ALLLLGVLLGLVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRDLLCFRA------ 65
Db    6 ALLLLMSLNLVHG-----VFY----TERYQTPTG-IKGPSNTKTQFFIPYAIKGKGVSLR 56

Qy   66 -RPPGP-----HSSPNYEFYKLYLVG----------GPGKPGTGSGPGPQCGPGLPGPPGS---A 104
Db   57 GEQGIPEGPPGPPAGPRGHPGPS------------------------ 103

Qy  104 CDRPDLDLRFTIKFQEYSPNLWGHEFRSHHDYYIIATSDGTREGLESLQG----- 153
Db  105 TGKPGL------PGLPGKQ------GER--GLNGPKGDIGPAGLPGP 137

Qy  154 ------GVCLTRGMKVLLRVG-QSPRGGAVPR-----KPVSEMP-----MERDRGAAHSL 196
Db  138 RGPPGPPGIPGPAGISVPGRKPGPQGPTGEPGPRGPPGPGEKGTSGVPGLNGQKGEMGHCTPC 197

Qy  197 EPGKENLPGDPTSNATSRGAEGEPLPPP------SMPAVAGAAGGLALLLLLGVAGAGGA 248
Db  198 RPGERGLPG-------PQGPPGPPGPPGVGKRGENGLPGQPGLKGD-----QGVPGERGA 245

Qy  249 MCWRRRAKPSESRHP----GPGSFGRGGSLGLGGGGGMGPREAEPGELGIA------ 296
Db  246 AGPSGPGPQGPGEQGPGEIGKPGAPGIPGQPGPGIPGMKQPGPGAGTAGLPGAPGF 298

Qy  297 ----LRG-GGAADPPFCPHYEKVSGDYGHPVIVQDGP-----PQSPPNI 336
Db  299 GKPGLPGLKGQRGPVGLPGSPGAKGEQGPAGHPGEAGLPGPSGNMGPQGPKGI 351
```

RESULT 28
S13580
collagen alpha 1(IX) chain precursor, long splice form - human
N;Alternate names: procollagen alpha 1(IX) chain, long splice form
C;Species: Homo sapiens (man)
C;Date: 13-Jan-1995 #sequence revision 10-Feb-1995 #text_change 20-Sep-1999
C;Accession: S13580; S23295; S02140; C35980; S21087; S74294
R;Muragaki, Y.; Kimura, T.; Ninomiya, Y.; Olsen, B.R.
Eur. J. Biochem. 192, 703-708, 1990
A;Title: The complete primary structure of two distinct forms of human alpha-1(IX) coll
A;Reference number: S13580; MUID:91006164; PMID:2209617
A;Accession: S13580
A;Molecule type: mRNA
A;Residues: 1-931 <MUR>
A;Cross-references: EMBL:X54412; NID:g30083; PIDN:CAA38276.1; PID:g30084
R;Ninomiya, Y.; Castagnola, P.; Gerecke, D.; Gordon, M.K.; Jacenko, O.; LuValle, P.; Mc
maguchi, N.; Olsen, B.R.
A;Title: The molecular biology of collagens with short triple-helical domains.
In Extracellular Matrix Genes, Sandell L.J. and Boyd C.D., eds., pp.79-114, Academic Pr
A;Reference number: S22243
A;Accession: S23295
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-253, 'V',255-815;835-884 <NIN>
C;Accession: S23295; S02140; C35980; S21087; S74294
R;Kimura, T.; Mattei, M.G.; Stevens, J.W.; Goldring, M.B.; Ninomiya, Y.; Olsen, B.R.
Eur. J. Biochem. 179, 71-78, 1989
A;Title: Molecular cloning of rat and human type IX collagen cDNA and localization of t
A;Reference number: S02140; MUID:89137096; PMID:2465149
A;Accession: S02140
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 580-596;598-620,'R',622-813;835-884 <KIM>
R;Muragaki, Y.; Nishimura, I.; Henney, A.; Ninomiya, Y.; Olsen, B.R.
Proc. Natl. Acad. Sci. U.S.A. 87, 2400-2404, 1990
A;Title: The alpha1(IX) collagen gene gives rise to two different transcripts in both m
A;Reference number: A35980; MUID:90207204; PMID:1690886
A;Accession: C35980
A;Molecule type: DNA
A;Residues: 1-4,233-267 <MUR2>
A;Cross-references: EMBL:M32135
submitted to the EMBL Data Library, February 1990
A;Reference number: S21087
A;Accession: S21087

A;Molecule type: DNA
A;Residues: 1-4,233-248,'T',250-267 <OLS>
A;Cross-references: EMBL;M32135
A;Map position: 6q12-6q14
A;Introns: 5/2; 232/3; 260/3; 267/3
C;Complex: type IX collagen may be a heterotrimer of one alpha 1(IX) chain, one alpha 2(
C;Function:
A;Description: structural component of extracellular fibrous polymer associated with typ
A;Note: in chondrocytes the long splice form is predominantly produced
C;Superfamily: unassigned collagens
C;Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; heter
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-268/Domain: non-collagenous NC4 #status predicted <NC4>
F;24-931/Product: collagen alpha 1(IX) chain, long splice form #status predicted <MAT>
F;269-405/Domain: collagenous COL3 #status predicted <COL3>
F;406-417/Domain: non-collagenous NC3 #status predicted <NC3>
F;418-756/Domain: collagenous COL2 #status predicted <COL2>
F;757-786/Domain: non-collagenous NC2 #status predicted <NC2>
F;787-901/Domain: collagenous COL1 #status predicted <COL1>
F;902-931/Domain: non-collagenous NC1 #status predicted <NC1>
F;171/Binding site: carbohydrate (Asn) (covalent) #status predicted

A;Molecule type: protein
A;Accession: S74294
A;Residues: 405-417 <DIA>
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit c
ed and subsequently O-glycosylated.
C;Genetics:
A;Gene: GDB:COL9A1
A;Cross-references: GDB:119794; OMIM:120210

R;Diab, M.; Wu, J.J.; Eyre, D.R.
Biochem. J. 314, 327-332, 1996
A;Title: Collagen type IX from human cartilage: a structural profile of intermolecular cr
A;Reference number: S64673; MUID:96195147; PMID:8660302

```
Query Match     5.9%;  Score 145;  DB 2;  Length 931;
Best Local Similarity  22.3%;  Pred. No. 0.015;
Matches   84;  Conservative  21;  Mismatches 120;  Indels 152;  Gaps 13;

Qy     2  GPPHSGPGGVRVGALILLGVILGVSGLSLEPVYWNSANKRFQAEGGYYLYPQIGDRLDLI   61
Db   391  GPP----GPGPR------------------------------------------GTIGFHDG   410

Qy    62  CPRARPPGPHSSPNVEFYKLYLVG-------GAQGRRCEAPPAPNLLLTCDRPDLDLRFTIK  116
Db   411  CPNACPGRSGYPG-------------------------LPGMRGHKGAKGEIGE--------  440

Qy   177  PRKPVSEMPERDRGAAHSL--EPGKENLPGDPTSNAT-----------SRGAEGPLPPP  223
Db   469  GLRGITGLVGDKGEKGARGLDGEPGPQGLPGAPGPDQGGPGEAGPKGDRGAEGARGIP  528

Qy   224  SMPAVAGNAAGGLALILLGVIAGAGGAMCWRRRAKPSESRHPG-------PGSFGRGGSLG  276
Db   529  GLPGVDGRDGIPGMPGTKGEPGKPGPGPGDAGLQGLPGVPGIPGAKG  583

Qy   277  LGGGGGMGPREAEPGELGTALRGG--------------GAADPFFCPHYEKVSG  316
Db   584  VAGEKGSTGAPGKFGQMGNSGKPGQQGPGEVGPRGPGPQGLPGSRGRGELGPVGSPGLPGKLG  643

Qy   317  DYGHGPVYIVQDGPPQSP  333
Db   644  SLGSPGLPGLPGPPGLP  660
```

RESULT 29
S53787
collagen alpha chain - Paralvinella grasslei (fragments)
C;Species: Paralvinella grasslei
C;Date: 19-Mar-1997 #sequence_revision 18-Jul-1997 #text_change 07-May-1999
C;Accession: S53787

R;Gaill, F.; Mann, K.; Wiedemann, H.; Engel, J.; Timpl, R.
J. Mol. Biol. 246, 284-294, 1995
A;Title: Structural comparison of cuticle and interstitial collagens from annelids livi
A;Reference number: S53786; MUID:95173973; PMID:7869380
A;Accession: S53787
A;Molecule type: protein
A;Residues: 1-90;51-254;255-304;305-374;375-438 <GAI>
C;Superfamily: unassigned collagens
C;Keywords: coiled coil; extracellular matrix; glycoprotein

```
Query Match     5.9%;  Score 144.5;  DB 2;  Length 438;
Best Local Similarity  26.5%;  Pred. No. 0.007;
Matches  101;  Conservative  15;  Mismatches 124;  Indels 141;  Gaps 22;

Qy     2  GPP-----HSGPGGVRVGALILLGVILGVGLVGS-------------GLSLEPVYWNSANKRFQ  43
Db   105  GPPGLTGSXGPAG--PXGANGLPGGTQGPGPAGEAGPXGSPGPGPGERGLGGHP---GEAGKRGQ  160

Qy    44  AEGGYVLYPQ-------IGDRLDLLCPRAPP------------GPHSSPNVEFYKLYLVG  85
Db   161  RGVPGALGPPGPGGERGLPGER---GLPGSRGPGGARGGPGERGPSGPP---------G  207

Qy    86  GAQG--RRCEAPPAPNLLLTCDRPDLDLRFTIKFQEYSPNLWGHEFRSHHDYYIIATSDG  143
Db   208  GPQGDGGRAGTPGDAPGLRGAPGRPG----------------------------SDG  235

Qy   144  TRBGLESLQG-----GVCLTRGMKVLLRVGQSPRGGAVPRKPVSEMPERDRGAAHSLEP  198
Db   236  -REGKPGSSGXPGKDG--------RPDGPRG---------------LPGRGAGS--N  269

Qy   199  GKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAG--GLALLLLGVIAGAGGAMCWRRRA  256
Db   270  GEDGQNGKP------------GGKGEVGPPGGTGLPLPGERGAKGELQGLPGGLPGEERGK---PGERG  319

Qy   257  KPSESRHPG-----PGSFGRGGSLGLGGGGGMGPREAEPGELGIALRGGGAADPPFCPHYE  312
Db   320  KDGESGHPGTAGLPGPGPQGV--KGERGDHGPR-GEPGSPG----SXGERGPPGXPXPQ  372

Qy   313  KVSGDYGHPVIVQDGPPQSP  333
Db   373  GLRGDRGNTGERGEPGAPGGP  393
```

RESULT 30
S28774
collagen alpha chain - tube worm (Riftia pachyptila) (fragment)
C;Species: Riftia pachyptila
C;Date: 22-Nov-1993 #sequence_revision 09-Mar-1996 #text_change 23-May-1997
C;Accession: S28774; S22915; S17581
R;Mann, K.; Gaill, F.; Timpl, R.
Eur. J. Biochem. 210, 839-847, 1992
A;Title: Amino-acid sequence and cell-adhesion activity of a fibril-forming collagen fro
A;Reference number: S28774; MUID:93130909; PMID:1483468
A;Molecule type: protein
A;Residues: 1-95;'X',97-107;'X',109-191;'X',193-260;'X',262-278;'X',280-572;'X',574-611,
A;Accession: S28774
A;Reference number: S22915
A;Accession: S22915
A;Description: Amino acid sequence and cell adhesion activity of a fibril-forming collag
A;Molecule type: protein
A;Residues: 903-proline modified to 4-hydroxyproline was also found
A;Note: we have shown the unidentified residues as Lys forming glycosylated 5-hydroxyly
R;Mann, K.; Gaill, F.; Timpl, R.
submitted to the Protein Sequence Database, July 1992
A;Reference number: S22915
A;Accession: S22915
R;Gaill, F.; Wiedemann, H.; Mann, K.; Kuehn, K.; Timpl, R.; Engel, J.
J. Mol. Biol. 221, 209-223, 1991
A;Title: Molecular characterization of cuticle and interstitial collagens from worms col
A;Reference number: S17581; MUID:92015209; PMID:1920405
A;Accession: S17581
A;Molecule type: protein
A;Residues: 8-45;'X',525-545,'X',547-566,'X',568-572,'X',574-611,'X',613-618,'X',811-882 <GA
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit (

h 3- and 4-hydroxylated in the X-position. Lysines are 5-hydroxylated and subsequently a

```
C;Complex: homotrimer
C;Superfamily: unassigned collagens
C;Keywords: cell binding; coiled coil; extracellular matrix; glycoprotein; homotrimer; h
C;Date:
F;1-12/Domain: amino-terminal telopeptide (fragment) <NTE>
F;1-12/Domain: collagenous telopeptide <COL>
F;13-1023/Domain: carboxyl-terminal telopeptide
F;1024-1027/Domain: carboxyl-terminal telopeptide (fragment) <CTE>
F;21,24,123,243,276,285,291,303,648,645/Modified site: 4-hydroxyproline (Pro
F;27,39,54,72,90,93,128,150,162,165,174,177,180,207,216,219,228,237,249,255,306,312,321,
F;711,714,717,723,744,759,774,783,792,816,843,849,855,861,867,888,894,915,945,954,963,966
F;53,161,165,416,551,647,773,815,1010,1011,1016,1019/Modified site: 3-hydroxyproline (Pr
F;96,108,192,261,279,573,612,657,738,765,810,927,936/Modified site: 5-hydroxylysine (Lys
F;96,108,192,261,279,573,612,657,738,765,810,927,936/Binding site: carbohydrate (Lys) (o
F;183,342,546,567,939/Modified site: 5-hydroxylysine (Lys) (partial)
F;351,933/Modified site: 5-hydroxylysine (Lys) (partial) #status experimental
```

Query Match

Best Local Similarity   22.2%;   Score 144; DB 2; Length 1027;
Matches   82; Conservative 24; Mismatches 113; Indels 150; Gaps 17;

```
Qy   46  GGYVLYFQIGDRLDLLCPRARPPGFHSSPNYEYKLIVLVGGAQGRRCE---------  93
             |.|    :.|.      ||| |   |: . |     | : |  |::
Db    4  GPRYIQAQVGP----IGPRG--PPGPPGSPGQQ--------GYQGLRGEPGDSGPMGPIGK  50

Qy   94  -APPAPNLLLTCDRPDLDIRFTIKFQEYSPNLWGHEFRSHHDYYIIATSDGTREGLESLQ  152
                 :    :  :|       :|:.         ||||:::|:|||
Db   51  RGPPGP--------------AGIAG------------KSGDDGRDGEPGR  75

Qy  153  GGVCLT-----------RGMKVL----------LRVGQSPRGG------AVPRKPVSEM  184
         |.| |           || : |          |:   : ||          :|   :|
Db   76  GGIGPMGPRGAGMPGMPGMPGKGHGRGRGLSGKGKEQGKSGNQGPDGPDGPPGPAGPSGPIGPR  135

Qy  185  PMERDRGAAHSLEPGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAGGIALL-----  238
              |||||:::|:|||      ::  : | |   | ::  |
Db  136  GQTGERG----RDGKSGLPG------LRGVDGLAGPPGPPGPIGSTGSPGFPGTPPGSK  183

Qy  239  -----------LLGVAGAGGAMCWRRRAKPSESRHPG-PGSFGRGGSLGLGGGGGM-----  283
                       :.| ||    ::  :|           ||:|:||| |: ||:|
Db  184  GDRGQSGIKGAQGLQGPVGLSGQPGVAGENGHPGMPGMDGANGEPGASGESGLPGPSGFP  243

Qy  284  GPR------EAAEPGELGIALRGG---------------GAADPPFCPHYEKVSGDYGHPVYI  324
         | :       |:|: | :|| ||                 | :| ||:||  :||:||||||
Db  244  GPRGMPGTAGSPGQAGARKGDGGPTGEQGQRGPGAPGVKGSSGGPP------GDVGAPGHA  294

Qy  325  VQDGPPQSP 333
         :|||| ||
Db  295  GEAGKRGSP 303
```

RESULT 31
A54121
collagen alpha-4 chain precursor - sea urchin (Strongylocentrotus purpuratus)
```
N;Alternate names: collagen alpha 2(IV) chain homolog
C;Species: Strongylocentrotus purpuratus (purple urchin)
C;Date: 07-Jul-1995 #sequence_revision 07-Jul-1995 #text_change 13-Aug-1999
C;Accession: A54121; A54317
R;Exposito, J.Y.; Suzuki, H.; Geourjon, C.; Garrone, R.; Solursh, M.; Ramirez, F.
J. Biol. Chem. 269, 13167-13171, 1994
A;Title: Identification of a cell lineage-specific gene coding for a sea urchin alpha2(I
A;Reference number: A54121; MUID:94230414; PMID:8175744
A;Accession: A54121
A;Molecule type: mRNA
A;Residues: 1-1747 <EXP>
A;Cross-references: EMBL:X76730; NID:g483606; PIDN:CAA54146.1; PID:g483607
C;Genetics:
A;Gene: COL4alpha
C;Superfamily: collagen alpha 1(IV) chain
```

Query Match

Best Local Similarity   24.1%;   Score 143; DB 2; Length 1747;
Matches   84; Conservative 24; Mismatches 159; Indels 82; Gaps 14;

```
Qy    7  GPGGRVRGALLLLGVLGLVSGLSLEPVYWNSANRFQAEGGYVLYFQIGDRLDLLCPRAR  66
```

RESULT 32
A57131
collagen alpha 2(VIII) chain - human (fragment)
```
C;Species: Homo sapiens (man)
C;Date: 03-Nov-1995 #sequence_revision 03-Nov-1995 #text_change 13-Aug-1999
C;Accession: A57131
R;Muragaki, Y.; Jacenko, O.; Apte, S.; Mattei, M.G.; Ninomiya, Y.; Olsen, B.R.
J. Biol. Chem. 266, 7721-7727, 1991
A;Title: The alpha2(VIII) collagen gene. A novel member of the short chain collagen far
A;Reference number: A57131; MUID:91210292; PMID:2019595
A;Accession: A57131
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-635 <MPR>
A;Cross-references: GB:M60832; NID:g171178; PIDN:AAA62822.1; PID:g171179
C;Gene: GDB:COL8A2
A;Cross-references: GDB:127812; OMIM:120252
A;Map position: 1p34.3-1p32.3
C;Superfamily: collagen alpha 1(VIII) chain
F;1-11/Domain: amino-terminal nonhelical (fragment) #status predicted <NC2>
F;12-468/Region: interrupted helical
F;469-635/Domain: carboxyl-terminal nonhelical (fragment) #status predicted <NC7>
F;508-634/Domain: complement C1q carboxyl-terminal homology <C1Q>
```

Query Match

Best Local Similarity   26.1%;   Score 142.5; DB 2; Length 635;
Matches   74; Conservative 17; Mismatches 85; Indels 107; Gaps 14;

```
Qy  147  GLESLQG------GVCLTRGMKVLLRVGQSPRGGAVP-------RKPVSEMPERDRGAA  193
Db  222  GLPGPQGPSGAKGEPGTRGPPGLI----GPTGYGMGPGLPGPKGDGRGPAGVPGLLGDRG--  275

Qy  194  HSLEPGKENLPGD--------PTSNATS-------RGAEGPLPPPSMPAVAGAAG  233
Db  276  ---EPGEDGPGEQPGQGLGGPPGLPGSAGLPGRRGPPGLRGEAGPPGGPGVPGIRGDQG  332

Qy  234  --GIALLLLGVAGAGGAMCWRRRRAKPSESRHPG--------------GSFGRG  272
Db  333  PSGGLA-----GKPGVPG----ERGLPGAHGPPGPTGPKGEPGFTGRPGGPGVAGALGQK  382

Qy  273  GSLGL--------GGGGGMGR-----EAAEPGELGIALRG---------GGAAD  304
Db  383  GDLGLPGQPGLRGPPGSGIPGLGPAGPIGPQGLPGPGEPGLPGPPGEGRAGEPGTAGPRG  442

Qy  305  PPFCPHYEKVSGDYGHPVYIVQDGPPQSPNIYISISVLEWP 347
Db  443  PPGVGSPGITGPPGLP------GPPGAFGAFDETGIAGLHLP 479
```

```
Db  266  GPRGMDG-----MKGATGEVGDLGSYGDYGEKGLPGYSGRGFPGNPGIPGLLGMNGEKGR 321
Db   67  PPGPHSSPNYEFYKLIVLVGGAQGRRCEAPPAPNLLLTCDRPDLDLRFTIKFQEBYSPNLWG 126
Db  322  -DGPQGRPGIDGYK--GPSGDYGMDGDPGPAPDI----------------------  352
Db  127  HEFRSHHDYYIIATSDGTRGLESLQGGVCLTRGMKVLLRVGQ----SPRGGAVPRKPVS 182
Db  353  -----------BEIVGVGPGDEDGPGNPGPRGPSGGATGLLPGDGPGPGCYPGPWGGPGPTG 405
Db  183  EMPMERDRGAAHSLEPGKENLPGDP-----TSNATSRGAEGPL-----PPPSMPAVAGAA 232
Db  406  SQGEBGRDGEKGEESGRPGIPGFGQEFTGERGNDGNFGERGEKGNMGESRPGPPGMCGSRGRQ 465
Db  233  GGLALILLGVAGAGGAMCWRRRRAKPSESRHPG------------PGSFGRGGSLGLGGGGGMGP 285
Db  466  G-----FMGQKGRGGRGP-----PGRAGPAGPARSGNGNQSFQGPGDTGNTGLKGMRGIGPALG 516
Db  286  REAAEPGELGIALRGGGAADPPFCPHYEK-VSGDYGHPVYIVQDGPPQSP 333
Db  517  RDGRPGSKG------ELGGIC--PFCPGIKGYPGDRGYP-----GDPGSP 554
```

RESULT 33
S23779
collagen alpha 1 (VIII) chain - mouse
C;Species: Mus musculus (house mouse)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C;Accession: S23779
R;Muragaki, Y.; Shiota, C.; Inoue, M.; Ooshima, A.; Olsen, B.R.; Ninomiya, Y.
R;Murakami, Y.;
Eur. J. Biochem. 207, 895-902, 1992
A;Title: Alpha-1(VIII)-collagen gene transcripts encode a short-chain collagen polypepti
A;Reference number: S23779; MUID:92362626; PMID:1499564
A;Accession: S23779
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-743 <MUR>
A;Cross-references: EMBL:X66976; NID:g50493; PIDN:CAA47387.1; PID:g1359953
C;Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology
F;616-742;Domain: complement C1q carboxyl-terminal homology <C1Q>

Query Match   5.8%; Score 142.5; DB 1; Length 743;
Best Local Similarity   21.3%; Pred. No. 0.017; DB 1;
Matches 101; Conservative   29; Mismatches 136; Indels 209; Gaps   21;

```
Qy   14 GALLLLGVGLVSGLSLEPVYWNSANKRFQAEGGYV----LYPQIGDRLDLLCPRARPPG   69
            ||| |||| ::  | |||:  ||||||||:
Db    6 GPLQLLGILFIIS------------LNSV--RLIQAGAYYGIKPLPFQIPPQIPQYQPLG   55

Qy   70 ---PH--------------SSPNYEFYKLY-----------------------L   83
                               :::   : :
Db   56 QQVPFMPLGKDGLSWGKEMPHMQYGXEYPYLPQYWKEIPVPRNGKEVVPRKGSKEVPLAS  115

Qy   84 VGGAQGRRCR-----APPAPNLLLTCDRPDLDLRFTIKFQEYSPNLWGHEFRSHDYYIIA  139
           |||      ||      ::  | ||:
Db  116 LRGEQGPRGEPGPRGPPGP------------PGLPGHGMPG------IK          146

Qy  140 TSDGT-------REGLESLQG--GVCLTRGMKVLLRVGQSPRGGAVPRKPVSEMPMERDR  190
          : :        |||   |       :  |
Db  147 GKPGPGGYPGIGKPGMPGMPGKPGAMGMPGAK-------GEIGPKGEIGPMGIPGPQ     196

Qy  191 G--AAHSL---EPGKENLPGDPTSNATSRGAGGPLPPSM-----------         225
          |   |      ||::| |||     |  |||||  |
Db  197 GPPGPHGLPGIGKPGGPGDLPGQPGAKG-ERGPKGPPGPPGLQGPRGEKGPGMLGLPGLKG 255

Qy  226 -PAVAGAAGGLALLLLGVAGAGAMCWRRRRAKPSESRHPG--------          265
Db  256 PPGMPGPGPGVGLPGVGKPGVTGPPGPGPQGPLGKGPGPPGGERGPQGLIGVPGVQGPPGMPGV 315

Qy  266 --PGGSFGRGSSLGLGGGGG-----------MGPRE                       287
Db  316 GKPGGPGSRGQPGPPGGKGEQGLPGLPGPGLPGVGKPGGPGPGKGHRGIGGVPGVLGPR-  374

Qy  288 AEPGELGIALRGGGAADPPPCHYEKVSGDYGHGPVYIVQDGP------PQSPP      334
Db  375 GEKGPIGAPGMGG------PPGEPGLPGIPGMGPPGAIGPPGPKGEGGVVGPQGPP     425
```

RESULT 34
CGHU2V
collagen alpha 2 (V) chain precursor - human
C;Species: Homo sapiens (man)
C;Date: 31-Jul-1989 #sequence revision 28-Jul-1995 #text change 31-Dec-2000
A;Molecule type: mRNA
A;Residues: 1-463 <WOO>
A;Cross-references: GB:J04478; NID:g179697; PIDN:AAA51859.1; PID:g179698
R;Woodbury, D.; Benson-Chanda, V.; Ramirez, F.
J. Biol. Chem. 264, 2735-2738, 1989
A;Title: Amino-terminal propeptide of human pro-alpha2(V) collagen conforms to the struc
A;Reference number: A31427; MUID:89123368; PMID:2914927
A;Accession: A31427
A;Molecule type: mRNA
A;Residues: 1-463 <WOO>
A;Cross-references: GB:J04478; NID:g179697; PIDN:AAA51859.1; PID:g179698
R;Greenspan, D.S.; Lee, S.T.; Lee, B.S.; Hoffman, G.G.
Gene Expr. 1, 29-39, 1991
A;Title: Homology between alpha2(V) and alpha1(III) collagen promoters and evidence for
A;Reference number: A54555; MUID:92314691; PMID:1820205
A;Accession: A54555
A;Molecule type: DNA
A;Residues: 1-32 <GRE>
A;Cross-references: GB:M58529; NID:g180834; PIDN:AAC41699.1; PID:g553235
R;Moradi-Ameli, M.; Rousseau, J.C.; Kleman, J.P.; Champliaud, M.F.; Boutillon, M.M.; Se
Eur. J. Biochem. 221, 987-995, 1994
A;Title: Diversity in the processing events at the N-terminus of type-V collagen.
A;Reference number: S43642; MUID:94237164; PMID:8181482
A;Accession: S43643
A;Molecule type: protein
A;Residues: 288-291,'X',293-294,'X',296-297,606,'X',608-617 <MOR>
R;Weil, D.; Bernard, M.; Gargano, S.; Ramirez, F.
Nucleic Acids Res. 15, 181-198, 1987
A;Title: The pro alpha 2 (V) collagen gene is evolutionarily related to the major fibril
A;Reference number: A25874; MUID:87146331; PMID:3029669
A;Accession: A25874
A;Molecule type: mRNA; DNA
A;Status: translated from GB/EMBL/DDBJ
R;Myers, J.C.; Loidl, H.R.; Stolle, C.A.; Seyer, J.M.
J. Biol. Chem. 260, 5533-5541, 1985
A;Title: Partial covalent structure of the human alpha 2 type V collagen chain.
A;Reference number: I55239; MUID:85182703; PMID:2985598
A;Accession: I55239
A;Molecule type: mRNA
A;Residues: 398-1496 <WEI>
A;Cross-references: GB:X04758; NID:g295588; PIDN:CAA28454.1; PID:g1340175
R;Emanuel, B.S.; Cannizzaro, L.A.; Seyer, J.M.; Myers, J.C.
Proc. Natl. Acad. Sci. U.S.A. 82, 3385-3389, 1985
A;Title: Human alpha 1(III) and alpha 2(V) procollagen genes are located on the long arm
A;Reference number: I59025; MUID:85216505; PMID:3858826
A;Accession: I59025
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1002-1226 <RE2>
A;Cross-references: GB:M10956; NID:g180427; PIDN:AAA52007.1; PID:g180428
A;Note: part of this sequence were determined by protein sequencing
R;Myers, J.C.; Loidl, H.R.; Seyer, J.M.; Dion, A.S.
J. Biol. Chem. 260, 11216-11222, 1985
A;Title: Complete primary structure of the human alpha-2 type V procollagen COOH-termina
A;Reference number: A25374; MUID:85289337; PMID:2411731
A;Accession: A25374
A;Molecule type: mRNA
A;Residues: 1227-1417,'T',1419-1437,'S',1439-1496 <MYE>
A;Cross-references: GB:M11718; NID:g180912; PIDN:AAA52058.1; PID:g180913
A;Experimental source: normal fibroblasts
R;Tsipouras, P.; Schwartz, R.C.; Liddell, A.C.; Salkeld, C.S.; Weil, D.; Ramirez, F.
Genomics 3, 275-277, 1988
A;Title: Genetic distance of two fibrillar collagen loci, COL3A1 and COL5A2, located on
A;Reference number: A30017; MUID:89138450; PMID:3224983
A;Accession: A30017
A;Molecule type: DNA
A;Residues: 1449-1463,'E',1465-1495,'A' <TSI>
A;Cross-references: GB:J03051; NID:g179695; PIDN:AAA51858.1; PID:g179696
A;Note: the authors translated the codon GAA for residue 1460 as Gln, and GAG for residu
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit (
are 5-hydroxylated and subsequently O-glycosylated.
C;Comment: The amino-terminal propeptide domain appears not to be completely cleaved.
C;Genetics:
C;Gene: GDB:COL5A2
C;Cross-references: GDB:119064; OMIM:120190
A;Map position: 2q31-2q31
A;Introns: 33/1, 812/3; 830/3; 848/3; 884/3; 902/3; 922/3; 974/3; 1046/3; 1064/3; 1448/3
C;Complex: type V collagen may be a homotrimer of alpha 1(V) chains (see PIR:CGHU1V), a
alpha 2(V) chain and one alpha 3(V) chain, initially linked by disulfide bonds among the
ength, is formed with desmosine cross-links made from lysine and allysine residues
A;Description: structural component of extracellular fibrous polymer associated with cel

A;Note: may play a role in controlling the lateral growth of collagen I fibrils
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C;Keyword: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxyprol:
F;1-26/Domain: signal sequence #status predicted <SIG>
F;27-1250/Product: collagen alpha 2(V) chain #status predicted <MAT>
F;27-193/Domain: amino-terminal propeptide (uncleaved) #status predicted <NPP>
F;40-99/Domain: von Willebrand factor type C repeat homology <VWC>
F;109-186/Region: helical
F;187-208/Region: nonhelical
F;209-1225/Region: helical
F;503-505/Region: cell attachment (R-G-D) motif
F;941-943/Region: cell attachment (R-G-D) motif
F;1064-1066/Region: cell attachment (R-G-D) motif
F;1067-1069/Region: cell attachment (R-G-D) motif
F;1097-1099/Region: cell attachment (R-G-D) motif
F;1124-1126/Region: cell attachment (R-G-D) motif
F;1133-1135/Region: cell attachment (R-G-D) motif
F;1225-1250/Region: carboxyl-terminal nonhelical telopeptide
F;1251-1496/Domain: fibrillar collagen carboxyl-terminal homology <FCC>
F;1269-1496/Domain: carboxyl-terminal propeptide #status predicted <CPP>
F;27/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted
F;193-194/Cleavage site: Ala-Gln (procollagen N-endopeptidase) #status predicted
F;194/Modified site: Pyrrolidone carboxylic acid (Gln) (in mature form) #status predi
F;201/Modified site: allysine (Lys) #status predicted
F;290,293,296,608,1007,1013,1028,1034/Modified site: 4-hydroxyproline (Pro) #st
F;299,1139/Modified site: 5-hydroxylysine (Lys) #status predicted
F;299,1139/Binding site: carbohydrate (Lys) (covalent) #status predicted
F;1025/Modified site: 5-hydroxylysine (Lys) #status experimental
F;1025/Binding site: carbohydrate (Lys) (covalent) #status experimental
F;1250-1251/Cleavage site: Glu-Asp (procollagen C-endopeptidase) #status predicted
F;1259,1397,1325/Disulfide bonds: interchain #status predicted
F;1293,1299,1325/Disulfide bonds: interchain #status predicted
F;1333-1494,1402-1447/Disulfide bonds: #status predicted

Query Match   5.8%;   Score 142.5;   DB 1;   Length 1496;
Best Local Similarity   29.1%; Pred. No. 0.038;
Matches 55; Conservative 15; Mismatches 76; Indels 43; Gaps 9;

Qy   169 QSPRG--GAV-PRKPVSEMPMERDRGAAHSLEPGKENLPG---------DPTSNATSRGAE 217
              :| || |: | : |::| |::||:| |:| || |::|:: |::
Db   500 RGPRGDPGTLGPPGPVGERGAPGPNGF------PGSDGLPGPKGAQGERGPVGSSGPKGSQ 554

Qy   218 GPLPPPSMPAVAGAGGLALLLLGVAGAGGAMCWRRRAKPSESRHPGP---GSFGRGG 273
              |:|:|| || |||| : |: :| | |||||||
Db   555 GDPGPRGPEGPGLPGARG-------LTGNPGVQGPBGKGLGPLGAPGEDGRPGPPGSIGIKGPG 609

Qy   274 SLGLGGGGWGPREAPGPELGIALRGGGAADPPFCPHYEKVSGDYGHPVYIVQDGPP--- 330
            : |: | : | ||| | :|:|:|| || | :|: |: |:| :|| ||
Db   610 TMGLPGPKGSNGDPGKPGPAG----NPGVPGQRGAPGKDGKVGPYGPP-----GPPGLR 659

Qy   331 -----QSPP 334
                 ||||
Db   660 GERGKQGPP 668

RESULT 35
T33149
hypothetical protein F29C4.8 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 31-Jan-2000
A;Accession: T33149
R;Bradshaw, H.; Clarke, K.
A;Reference number: Z21291
A;Description: The sequence of C. elegans cosmid F29C4.
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-614 <BRA>
A;Cross-references: EMBL:AF076616; PIDN:AAC19194.1; GSPDB:GN00022; CESP:F29C4.8
A;Experimental source: strain Bristol N2; clone F29C4
C;Genetics:
A;Gene: CESP:F29C4.8

Query Match   5.8%;   Score 142;   DB 2;   Length 614;
Best Local Similarity   23.5%; Pred. No. 0.019;
Matches 73; Conservative 16; Mismatches 96; Indels 126; Gaps 14;

Qy   67 PGPHSSPNYEFYKLIVLVGGAGGRRCEAPPAPNLLLTCDRPDLDLRFTIKFQEYSPNLWG 126
             |||| : | | | | : ||::: :   :: : :: :|| |
Db   198 PPGPHGLPG------IGKPGPGLPGQPGPK------------------------------ 222

Qy   127 HEFRSHHDYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKPVSEMPM 186
            : |: |||: |:| | : | |:   :: |: |||:||: |
Db   223 ----------GDRGPKGLPGPQG-----------LRGPKGDKGFGPMGPAPGGVKGP- 256

Qy   187 ERDRGAAHSLEPGKENLPGDPTSNATS-RGAEGPLPPPSMPAVAGAGGLALLLL----G 241

RESULT 36
S15435
collagen alpha 1(VIII) chain precursor - human
C;Species: Homo sapiens (man)
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 13-Aug-1999
A;Accession: S15435
R;Muragaki, Y.; Mattei, M.G.; Yamaguchi, N.; Olsen, B.R.; Ninomiya, Y.
Eur. J. Biochem. 197, 615-622, 1991
A;Title: The complete primary structure of the human alpha-1(VIII) chain and assignment
A;Reference number: S15435; MUID:91211001; PMID:2029894
A;Accession: S15435
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-744 <MUR>
A;Cross-references: GDB:128104; OMIM:120251
A;Cross-references: EMBL:X57527; NID:g30081; PIDN:CAA40748.1; PID:g30082
C;Genetics:
A;Gene: GDB:COL8A1
A;Map position: 3q11.1-3q13.2
C;Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology <C1Q>
F;1-20/Domain: signal sequence #status predicted <SIG>
F;21-744/Product: collagen alpha 1(VIII) chain #status predicted <MAT>
F;21-117/Region: amino-terminal nonhelical
F;118-571/Region: interrupted helical
F;572-744/Region: carboxyl-terminal nonhelical
F;617-743/Domain: complement C1q carboxyl-terminal homology <C1Q>

Query Match   5.8%;   Score 142;   DB 2;   Length 744;
Best Local Similarity   27.2%; Pred. No. 0.015;
Matches 87; Conservative 19; Mismatches 110; Indels 104; Gaps 21;

Qy   57 RLDLLCPRARPP---GPHSSPNYEFYKLIVLVGGAQGGRRCEAPPAPNLLLTCDRPDLLRF 113
             | ||:| || ::| :|: |:| | || | | : | ||:: : | |
Db   24 RNSCMCP-AGPPGERGPVGPP-----------GLPGL----PAPYY---RRPRVPL-- 60

Qy   114 TIKFQEYSPNIWGHEFRSHHDYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQSPRG 173
            :| :|::   :| | : |:|: |:|| : || : | | | | | | |
Db   61 ------SNNLDBSISRKMRAFGMLYSPDG-----QAIQ-----LRGM-----PGPPG 96

Qy   174 GAVPR----KPVSEMPMERDRGAAHSLEPGKENLPGDPTSNATSRGAEGPLPPPSMPAVA 229
              | |: ||:| |::| :: ||| | |||||:||:|:|::|:|::|:||:|| :|
Db   97 PAGPKGLRGYPGPGPIGLD------GPRGLPGTPGPGSKG-DRGERGPPGPPGPFPGPGPK 146

Qy   230 GAAGGLALLLLGVAGAGGAMCWRRRAK--PSESRHPG-PGSFGRGGSLGLGGGGGM--- 283
            | ||| :| | | | | | | |: | || | | :|:| | |
Db   147 GDRGVMTGPYVGPHAGPGPMSHHTNMGNVLPGPPGPPGPGPAGRDGRHGLKGDRGLPGF 206

Qy   284 -----GPREARPGELGIALRGG--GAADPP------------FCPHYEKVSGDY----GH 320
                 | : | :| ::|:| :||| | |::|: | |:
Db   207 DGESKIGPK---GETGSPGRDGIPGARGPPGGERGEKGDTAFLSTYPRVASSSTASSPGP 262

Qy   321 PVYIVQDGPPGPSPPNIYTS 340
             : :: | || || |
Db   263 P-------GPP-GPPGVCHAS 275

```
Db  257  -----PGMHGL-PGPVGLPGVGRGPGGVTGPGPQGPLGKGPGAPGEPGRGQPGIGVPGVQGPG  311
Qy  242  VAGAGGAMGWRRRAKPSSRHPG------------------------PGSFGRGGS  274
Db  312  IPGIG--------KPGQDGIPGGQGPFGGKGEQGLGPLPGARGLPGIGKGFGPPGKGD  361
Qy  275  LGLGG-GGGMGPGREAEPGELGIALRGGGAADP--PFCPHYEKVSGDYGHPVVIVQDGP--  329
Db  362  RGMGGVPGALGPR-GEKGPIGSPGIGGSPGEPGLPGIP------GPMGPPGAIGFPGPGEKG  414
Qy  330  ------PQSPP  334
Db  415  EGGIVGPGGPP  425
```

RESULT 37

S21369
collagen alpha 2(VI) chain precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 19-Jan-2001
C;Accession: S21369; S28808; S13745; S18863
R;Ibrahimi, A.; Bertrand, B.; Bardon, S.; Amri, E.Z.; Grimaldi, P.; Ailhaud, G.; Dani, C.
Biochem. J. 289, 141-147, 1993
A;Title: Cloning of alpha2 chain of type VI collagen and expression during mouse develop
A;Reference number: S28808; MUID:93143559; PMID:8380980
A;Accession: S28808
A;Molecule type: mRNA
A;Residues: 1-1029 <IBR>
A;Cross-references: EMBL:X65582; NID:g49808; PIDN:CAA46541.1; PID:g49809
R;Ibrahimi, A.; Bardon, S.; Anri, E.Z.; Grimaldi, P.
Constantinou, C.D.; Jimenez, S.A.
Matrix 11, 1-9, 1991
A;Title: Structure of cDNAs encoding the triple-helical domain of murine alpha-2(VI) col
rate oligonucleotides for generation of novel cDNA clones.
A;Keywords: alternative splicing; cartilage; cell binding; connective tissue; disulfide
A;Reference number: S13745; MUID:9126374; PMID:1709252
A;Accession: S13745
A;Molecule type: mRNA
A;Residues: 266-1029 <IB2>
A;Cross-references: EMBL:X62332; NID:g192671; PIDN:CAA44206.1; PID:g49907
A;Molecule type: mRNA
A;Residues: 266-267 <IB>
A;Cross-references: GB:L06343; NID:g192671; PIDN:AAA37441.1; PID:g192672
A;Note: the sequence from Fig. 3 is inconsistent with that from Fig. 4 in having 306-Ile
A;Superfamily: collagen alpha 1(VI) chain; von Willebrand factor type A repeat homology
C;Keywords: alternative splicing; cartilage; cell binding; connective tissue; disulfide
F;1-28/Domain: signal sequence #status predicted <SIG>
F;29-1029/Product: collagen alpha 2(VI) chain #status predicted <MAT>
F;29-265/Domain: globular #status predicted <NC2>
F;54-226/Domain: von Willebrand factor type A repeat homology <VWA>
F;266-600/Domain: collagenous #status predicted <COL>
F;376-378/Region: cell attachment (R-G-D) motif
F;436-438/Region: cell attachment (R-G-D) novel
F;499-501/Region: cell attachment (R-G-D) motif
F;508-510/Region: cell attachment (R-G-D) motif
F;549-551/Region: cell attachment (R-G-D) motif
F;601-1029/Domain: globular #status predicted <NC1>
F;623-799/Domain: von Willebrand factor type A repeat homology <VWA2>
F;841-1012/Domain: von Willebrand factor type A repeat homology <VWA3>
F;150,337,640,907,963/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match   5.8%; Score 142; DB 1; Length 1029;
Best Local Similarity 29.7%; Pred. No. 0.027;
Matches   90; Conservative   16; Mismatches 125; Indels 72; Gaps 20;

```
Qy   68  PGPHSSPNYEFYKLYLVGGAQGRRCEAPPAPNLLITCDRPDLDRFTIKFQEYSPNLWGH  127
Db  265  PGPHGPGKGYRGQK----GAKGNMGR-PGEPGQKGRQGDPGIEGPIGPGPGKGVPGPKGE  318
Qy  128  --EFRSHHDYIIATSDGTREGLESLQG--GVCLTRGMKVLL-RVG-QSPRGGAVPRKP-  180
```

```
Db  319  KGEP---------GSDG-RKGAPGLAGKNG----TDGQKGKLGRIGPPGPGKGDPGSRGPD  364
Qy  181  -----VSEMPMER-DRGA-AHSLEPGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAGG  234
Db  365  GYPGEAGSPGERGDQGAAKGDSGRPGRRGPGPGPGDP-GDKGKSKGYQGNNGAPGSPGVKGKGG  423
Qy  235  LALLLGVAGAGAMCWRRRAKP------------------SESRHPGP----GSFGRGGSLGLG  278
Db  424  ------PGPRGPKGEPGRRGDPGTKGPSGSDPGKGEKGDPGEPGRGLAGEVGSKGAK  475
Qy  279  GGGGM-GPREAEPGELGIALRGGGAADPPFCPHYEKVSGDYGHPVVIVQDGPPQSP--PN  335
Db  476  GDRGLPGRPGRPQ-GALGEPGKQGSRGDP-------------GDAGPRGDSGQPGPKGDPRPG  524
Qy  336  IYY  338
Db  525  FSY  527
```

RESULT 38

S16366
collagen alpha 2(IV) chain precursor - pig roundworm
C;Species: Ascaris suum (pig roundworm)
C;Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 13-Aug-1999
C;Accession: S16366
R;Pettitt, J.; Kingston, I.B.
J. Biol. Chem. 266, 16149-16156, 1991
A;Title: The complete primary structure of a nematode alpha-2(IV) collagen and the part:
A;Reference number: S16366; MUID:91340768; PMID:1714907
A;Accession: S16366
A;Molecule type: mRNA
A;Residues: 1-1763 <JBI>
A;Cross-references: GB:M67507; NID:g159648; PIDN:AAA18014.1; PID:g159649
C;Genetics:
C;Introns: 229/3; 266/3; 305/3; 360/3; 424/1; 489/1; 548/1; 656/3; 790/1; 891/1; 963/1;
C;Superfamily: collagen alpha 1(IV) chain
C;Keywords: alternative splicing; basement membrane; cell binding; coiled coil; disulfi
F;1-26/Domain: signal sequence #status predicted <SIG>
F;27-1763/Product: collagen alpha 2(IV) chain #status predicted <MAT>
F;27-42/Domain: non-collagenous NH1 #status predicted <NH1>
F;43-1529/Domain: collagenous #status predicted <COL>
F;197-199/Region: cell attachment (R-G-D) motif
F;1530-1763/Domain: carboxyl-terminal nonhelical, NC1 #status predicted <NC1>
F;1530-1638/Domain: repeat NC1 #status predicted <NC11>
F;1639-1763/Domain: repeat NC1 #status predicted <NC12>
F;31,34,39,41,536,539/Disulfide bonds: interchain #status predicted
F;126/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;1593-1599,1702-1709/Disulfide bonds: #status predicted

Query Match   5.8%; Score 142; DB 2; Length 1763;
Best Local Similarity 25.5%; Pred. No. 0.049;
Matches   97; Conservative   24; Mismatches 127; Indels 132; Gaps 23;

```
Qy    2  GPPHSGPGGVRVGALLLLLGVLGL------VSGLSLEP----VYWNSANKRFQAEGGYVLY  51
Db   82  GPP---GPQGIK-GDRGIIGVPGFPGNDGANGRPGEPGPPGAPGWDGCNGTDGAPG---V  134
Qy   52  PQIGDRLDLLCPRARP--PGPHSSPNYEFYKLYLVGGAQGRRCEA--PPAPNLLITCDRP  107
Db  135  P-------GLPGPPGMPGFPGPPGVPGMKGEPAIGYAGAPGEKGDAGMPGMPGL------  181
Qy  108  DLDLRFTIKFQEYSPNLWGHEFRSHHDYIIATSDGTRGGLESLQGGVCLTRGMKVLLRV  167
Db  182  ------------PGPPG----------GAV-----PRKPVSEMPMERDRGAAHSLE--P  198
Qy  168  GQS-PRG------------------------------RDGFPGEKG----DRG----DV  200
Db  201  GQAGPRGPPGEAGPPGNPGIGSIGPKGDPGEQGPRGPQGPGPGPVPSTGAKGTIIGPPGAP  260
Qy  201  GPHGPGKGYRGQK----GAKGNMGR-PGEPGQKGRQGDPGIEGPIGPGPGKGVPGPKGE  318
Db  261  GMKGEKGDP-------------LPGMKGEKGLSGPAGPGKGE  308
```

```
Qy    259  SESRHPGPSFGRGGSLGLGGGGMGRPREAEPGELGIALRGG--GRAADPFPFGPHYEKVSG 316
Db    309  GRPGLPGPPGF--KGDRGLDLGPGVPGLPGQKQEAGPFPGRDGAKGARGPGPGPG----GG 362

Qy    317  DYGHPVYIVQDGPPQSPPNI 336
           ::
Db    363  EF-------SDGPP-GPPGL 374

Qy    317  DYGHPVYIVQDGPPQSPPNI 336
Db    363  EF-------SDGPP-GPPGL 374
```

RESULT 39

CGHU7L
collagen alpha 1(III) chain precursor - human
N;Alternate names: procollagen alpha 1(III) chain
C;Species: Homo sapiens (man)
A;Molecule type: mRNA
A;Status: preliminary
A;Accession: S05272
A;Reference number: S05272
A;Title: Structure of cDNA clones coding for the entire prepro-alpha1(III) chain of huma
erences.
A;Molecule type: mRNA
A;Residues: 1-1240,'V',1242-1466 <PRC>
A;Cross-references: EMBL:X14420; NID:g30057; PIDN:CAA32583.1; PID:g30058
R;Ala-Kokko, L.; Kontusaari, S.; Baldwin, C.T.; Kuivaniemi, H.; Prockop, D.J.
Biochem. J. 260, 509-516, 1989
R;Prockop, D.J.
submitted to the EMBL Data Library, February 1989
A;Reference number: S05272
A;Accession: S04642
A;Molecule type: mRNA
A;Reference number: S04642; MUID:89350838; PMID:2764886
A;Accession: S04642
A;Molecule type: mRNA
A;Residues: 1-1196 <ALA>
A;Cross-references: EMBL:X14420; NID:g30057; PIDN:CAA32583.1; PID:g30058
A;Note: the complete sequence is not shown
R;Benson-Chanda, V.; Su, M.W.; Weil, D.; Chu, M.L.; Ramirez, F.
Gene 78, 255-265, 1989
A;Title: Cloning and analysis of the 5' portion of the human type-III procollagen gene (
A;Reference number: PE0011; MUID:89378752; PMID:2777083
A;Accession: PE0011
A;Molecule type: DNA
A;Residues: 1-176 <BEN>
A;Cross-references: GB:M26939; NID:g180813; PIDN:AAA52040.1; PID:g180814
R;Toman, P.D.; Ricca, G.A.; de Crombrugghe, B.
Nucleic Acids Res. 16, 7201, 1988
A;Title: Nucleotide sequence of a cDNA coding for the amino-terminal region of human pre
A;Reference number: S01726; MUID:88303360; PMID:3405773
A;Accession: S01726
A;Molecule type: mRNA
A;Residues: 1-170 <TOM>
A;Cross-references: EMBL:X07240; NID:g30060; PIDN:CAA30229.1; PID:g30061
A;Note: the authors translated the codon CAG for residue 154 as His
R;Janeczko, R.A.; Ramirez, F.
Nucleic Acids Res. 17, 6742, 1989
A;Title: Nucleotide and amino acid sequences of the entire human alpha-1 (III) collagen.
A;Reference number: S04887; MUID:89386015; PMID:2780304
A;Accession: S04887
A;Molecule type: mRNA
A;Residues: 149-163,'G',164-240,'D',242-471,'L',473-487,'L',489,'S',491-613,'Y',615-634,'
A;Note: the authors' translation of residues 905-932 is inconsistent with the nucleotide
R;Kang, A.H.
Biochemistry 16, 1158-1164, 1977
A;Title: Covalent structure of collagen: amino acid sequence of cyanogen bromide peptide
A;Reference number: A90399; MUID:77134724; PMID:557335
A;Accession: A90399
A;Molecule type: protein
A;Residues: 'V',169-225,229-232,'P',234-292,'D',294-398 <SEY1>
R;Seyer, J.M.; Kang, A.H.
submitted to the Atlas, December 1977
A;Reference number: A94562

A;Accession: A94562
A;Molecule type: protein
A;Residues: 'V',169-225,229-277,'A',279-292,'D',294,'S',296-398 <SEY2>
A;Experimental source: liver
A;Note: author submitted corrections to A90399
R;Milewicz, D.M.; Witz, A.M.; Smith, A.C.; Manchester, D.K.; Waldstein, G.; Byers, P.H.
Am. J. Hum. Genet. 53, 62-70, 1993
A;Title: Parental somatic and germ-line mosaicism for a multiexon deletion with unusual
splicing.
A;Reference number: I51868; MUID:93304430; PMID:8317500
A;Accession: I51868
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 186-194 <ML>
A;Cross-references: GB:S62925; NID:g386425; PIDN:AAD13937.1; PID:g94261637
R;Chiodo, A.A.; Sillence, D.O.; Cole, W.G.; Bateman, J.F.
Biochem. J. 311, 939-943, 1995
A;Title: Abnormal type III collagen produced by an exon-17-skipping mutation of the COL
A;Reference number: S59511; MUID:96067614; PMID:7487954
A;Accession: A90414
A;Molecule type: protein
A;Residues: 399-675,'N',677-727 <SEY3>
A;Experimental source: liver
A;Cross-references: GB:S79877; NID:g1195576; PIDN:AAB35615.1; PID:g1195577
R;Seyer, J.M.; Kang, A.H.
J. Biol. Chem. 266, 5256-5259, 1991
R;Lee, B.; Vitale, E.; Superti-Furga, A.; Steinmann, B.; Ramirez, F.
J. Biol. Chem. 266, 5256-5259, 1991
A;Title: G to T transversion at position +5 of a splice donor site causes skipping of t
A;Reference number: I55349; MUID:91161621; PMID:1672129
A;Accession: I55349
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 537-605 <LEE>
A;Cross-references: GB:M59312; NID:g180815; PIDN:AAA52041.1; PID:g180816
R;Cole, W.G.; Chiodo, A.A.; Lamande, S.R.; Janeczko, R.; Ramirez, F.; Dahl, H.H.M.; Cha
J. Biol. Chem. 265, 17070-17077, 1990
A;Title: A base substitution at a splice site in the COL3A1 gene causes exon skipping a
A;Reference number: A38303; MUID:91009133; PMID:2145268
A;Accession: A38303
A;Molecule type: protein
A;Residues: 728-895,'A',897-964 <SEY4>
A;Experimental source: liver
A;Molecule type: DNA
A;Status: translation not shown
A;Reference number: S02119; MUID:88189827; PMID:3357782
A;Accession: S02119
A;Molecule type: mRNA
A;Residues: 861-1015 <COL>
A;Cross-references: GB:J05617; GB:M55603; GB:M59227; NID:g180878; PIDN:AAB59383.1; PID:
R;Mankoo, B.S.; Dalgleish, R.
Nucleic Acids Res. 16, 2337, 1988
A;Title: Human pro alpha1(III) collagen: cDNA sequence for the 3' end.
A;Reference number: A90446; MUID:81208139; PMID:7016180
A;Accession: A90446
A;Molecule type: protein
A;Residues: 950-1018,'Y',1020-1183,'S',1185-1466 <MAN>
A;Cross-references: EMBL:X06700; NID:g30053; PIDN:CAA29886.1; PID:g30054
R;Seyer, J.M.; Kang, A.H.
Biochemistry 20, 2621-2627, 1981
A;Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB9 from t
A;Reference number: A90446
A;Molecule type: protein
A;Residues: 965-979,'A',981-984,'PS',987,'QN',990-1096,'P',1098-1152,'AT',1155,'S',1157
```

A;Experimental source: liver
R;Loidl, H.R.; Brinker, J.M.; May, M.; Pihlajaniemi, T.; Morrow, S.; Rosenbloom, J.; Mye
Nucleic Acids Res. 12, 9383-9394, 1984
A;Title: Molecular cloning and carboxyl-propeptide analysis of human type III procollage
A;Reference number: A93551; MUID:85087944; PMID:6096827
A;Accession: A93551
A;Molecule type: mRNA
A;Cross-references: EMBL:X01655; EMBL:X01742; NID:g29584; PIDN:CAA25821.1
A;Cross-references: 1105-1155,'P',1157-1466 <LOI>
A;Residues: 1105-1155,'P',1157-1466 <LOI>
R;Miskulin, M.; Dalgleish, R.; Kluve-Beckerman, B.; Rennard, S.I.; Tolstoshev, P.; Brant
Biochemistry 25, 1408-1413, 1986
A;Title: Human type III collagen gene expression is coordinately modulated with the type
A;Reference number: 152393; MUID:86187804; PMID:3754462
A;Accession: 152393
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1161-1200 <MIS>
A;Residues: 1161-1200
A;Cross-references: GB:M13146; NID:g180415; PIDN:AAA52003.1; PID:g180416
R;Emanuel, B.S.; Cannizzaro, L.A.; Seyer, J.M.; Myers, J.C.
Proc. Natl. Acad. Sci. U.S.A. 82, 3385-3389, 1985
A;Title: Human alpha 1(III) and alpha 2(V) procollagen genes are located on the long arm
A;Reference number: 159025; MUID:85216505; PMID:3858826
A;Accession: I79359
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1165-1196 <EMA>
A;Cross-references: GB:M11134; NID:g180417; PIDN:AAA52004.1; PID:g180418
R;Chu, M.L.; Weil, D.; de Wet, W.; Bernard, M.; Sippola, M.; Ramirez, F.
J. Biol. Chem. 260, 4357-4363, 1985
A;Title: Isolation of cDNA and genomic clones encoding human pro-alpha1(III) collagen.
A;Reference number: A92516; MUID:85157600; PMID:2579949
A;Accession: A92516
A;Molecule type: DNA
A;Residues: 1176-1240,'V',1242-1356,'P',1358-1466 <CHU>
A;Residues: 1176-1240
A;Cross-references: GB:M10615; GB:M10793; GB:M10794; GB:M10795; GB:M10796; GB:M10797; GB
A;Experimental source: liver
A;Note: the authors translated the codon TTC for residue 1057 as Tyr; the codons given f
ation
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit (
3-hydroxylated. About 15% of the lysines are 5-hydroxylated and some are subsequently O
C;Genetics:
A;Gene: GDB:COL3A1
A;Cross-references: GDB:118729; OMIM:120180
A;Map position: 2q31-2q31
A;Introns: 27/1; 94/3; 111/3; 149/3; 176/3; 554/3; 587/3; 1175/3; 1275/1; 1337/3; 1418/3
A;Note: the list of introns is incomplete; defects in this gene can result in Ehlers-Dan
C;Complex: type III collagen is a homotrimer of monomers initially linked by disulfide b
er of their length, is formed with desmosine cross-links made from lysine and allysine r
C;Function:
A;Description: structural component of extracellular fibrous polymer that maintains inte
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C;Keywords: coiled coil; Ehlers-Danlos syndrome; extracellular matrix; glycoprotein; hyd
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-153/Domain: amino-terminal propeptide #status predicted <PRO>
F;31-91/Domain: von Willebrand factor type C repeat homology <VWC>
F;154-1221/Product: collagen alpha 1(III) chain #status predicted <MAT>
F;154-167/Region: amino-terminal nonhelical telopeptide
F;168-1196/Region: helical
F;1091-1093/Region: cell attachment (R-G-D) motif
F;1091-1093/Region: cell attachment (R-G-D) motif
F;1197-1221/Region: carboxyl-terminal nonhelical telopeptide
F;1222-1466/Domain: carboxyl-terminal propeptide #status predicted <CPR>
F;1238-1466/Domain: fibrillar collagen carboxyl-terminal homology <FCC>
F;24/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted
F;153-154/Cleavage site: Pro-Gln (procollagen N-endopeptidase) #status predicted
F;154/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicte
F;161/1212/Modified site: allysine (Lys) #status predicted
F;263,284,860,977,1106/Modified site: 5-hydroxylysine (Lys) #status experimental
F;263/Binding site: carbohydrate (Lys) (covalent) #status experimental
F;584,1094/Modified site: 5-hydroxylysine (Lys) (partial) #status experimental
F;949-949/Cleavage site: Gly-Ile (collagenase) #status experimental
F;1106/Binding site: carbohydrate (Lys) (covalent) #status predicted

Query Match   5.8%;   Score 141.5;   DB 1;   Length 1466;

RESULT 40
T32250
hypothetical protein T15B7.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 21-Jan-2000
R;Pauley, A.; Gattung, S.
submitted to the EMBL Data Library, September 1997
A;Description: The sequence of C. elegans cosmid T15B7.
A;Reference number: Z21139
A;Accession: T32250
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-319 <PAU>
A;Cross-references: EMBL:AF022985; PIDN:AAB69959.1; GSPDB:GN000023; CESP:T15B7.3
A;Experimental source: strain Bristol N2; clone T15B7
C;Genetics:
A;Gene: CESP:T15B7.3
A;Map position: 5
A;Introns: 266/1
C;Superfamily: unassigned collagens

Query Match   5.8%;   Score 141;   DB 2;   Length 319;
Best Local Similarity 25.6%;   Pred. No. 0.0087;
Matches 90;   Conservative 27;   Mismatches 111;   Indels 124;   Gaps 23;

Best Local Similarity 27.6%;   Pred. No. 0.043;
Matches 61;   Conservative 13;   Mismatches 74;   Indels 73;   Gaps 10;

Qy   13 VGALLLLGVLGLVSGLSL-------EPVYWNSANKRFQAEGG-----YVLYPQIGDRLDLLC 62
           :|| |||:|: ||:|: ||::||: |:   :|| |||:|:|::
Db   1 MSASTLVTVASRASGIAIVCVFTVGMIFNDINSFYDEKIGELKEFKGYEQIA--WQAMI 58

Qy   171 PRGGAVPRKPVSEMPMERDRGA-------AHSLEPGKENLPGDPTSNAT-----SRGAEG 218
Db   844 PPGGSGPAGPPGPGGVKGERGSPGGPGAAGFPGARGLPGPGPGSNGNGPPGPGSGSPGKDG 903

Qy   219 PLPP------------PSMPAVAGAAGGL-----------------ALLLLGVAGAGGAMCWRRRA 256
Db   904 PPGPAGNTGAPGSPGVSGGPKGDAGQPGEKGSPGGAQGPPGGARGPLGIAGITGARGLAGPPG 963

Qy   257 KPSESRHPGP----GSFGRGGSLGLGG-GGGMGP---------------REAAPGELGIAL 297
Db   964 MPGPRGSPGPQGGVKGESGKPGANGLSGERGPPGPGPGLPGLAGTAGEPGRDGNPGSDGLPG 1023

Qy   298 RGG--------GAADPPFCFHYEKVSGDYGHFVYIVQDGPP 330
Db 1024 RDGSPGGKGDRGENGSPGAP-------GAPGHP------GPP 1052

Qy   63 PRARPPGPHSSSNYEFYKLYLVG--------------GAQGRRCRA--PPAPNLLLTCDRDLDL 111
Db   59 PTTRPRSGGSS----------FLLGRNKRQAECNCGEQSRGCPAGPPGP-----GQPG--- 101

Qy   112 RFTIKFQEYSPNIWGHEFRSHHDYYIIATSDGTRGGLESLCQGGVCLTRGMKVLLRVGQSP 171
Db   102 ---ARGEAGLPGIAGQ--------PGSGARINPATGRPGPCIT--------------CP 135

Qy   172 RGGAVPRKPVSEMPMERDRGAAHSLEPGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGA 231
Db   136 AGAPGPAGP----------PGA----PGPKGNNGQPGAPAQS-GGRGP-PGPRGP-AGD 177

Qy   232 AGGLALLLLGVAGAGGAMCWRRRARAKPSESRHPG-PGSFGRGG--SLGLGGGGG----MG 284
Db   178 AG-------------SPGQPGHPGSPGSFGNPGRGGQRSRGTPGASGRPGPQG 214

Qy   285 PREAAPGELGIALRGGGAADP-PFCFHYEKVSGDYGHPVYIVQDGPPGPQSPPN 335
Db   215 PAGA-PGQPG--RSGGAGTPGPGGP------PGPSGQPGHSGNDGVPGTPGN 257

RESULT 41
A34246

collagen alpha 1(VIII) chain precursor - rabbit
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: A34246
R;Yamaguchi, N.; Benya, P.D.; van der Rest, M.; Ninomiya, Y.
J. Biol. Chem. 264, 16022-16029, 1989
A;Title: The cloning and sequencing of alpha1(VIII) collagen cDNAs demonstrate that type
omains similar to those of type X collagen.
A;Reference number: A34246; MUID:89380199; PMID:2476437
A;Accession: A34246
A;Molecule type: mRNA
A;Residues: 1-744 <YAM>
A;Cross-references: GB;J05042; NID:g164895; PIDN:AAA31204.1; PID:g164896
A;Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology
F;1-20/Domain: signal sequence #status predicted <SIG>
F;21-744/Product: collagen alpha 1(VIII) chain #status predicted <MAT>
F;21-117/Region: amino-terminal nonhelical
F;118-571/Region: interrupted helical
F;572-744/Region: carboxyl-terminal nonhelical
F;617-743/Domain: complement C1q carboxyl-terminal homology <C1Q>

Query Match   5.8%; Score 141; DB 1; Length 744;
Best Local Similarity   23.6%; Pred. No. 0.022;
Matches   83; Conservative   29; Mismatches 108; Indels 132; Gaps 19;

Qy   67 PPGPHSSPNYEFYKLYLVG----------------GAGQ-RRCEAPPAPNLLLTCDRPDLDLRFT 114
             |||| |: |||| : : | | | |
Db   198 PPGPHGLPG----------------IGKPGGPGLPGGPGAKGDRGPKGPGP--------------- 233

Qy   115 IKFQEYSPNLWGHEFRSHHDYYIIATSDGTR----------EGLESLQG--GVCLTRGMKVLLRVG 168
          | |||:||:|| ||| :|| ||||||| | | | :|| ||| | |
Db   234 ------------PGLQGPK-----------------GEKGFGMPGLPGLKGPPGMHGPPGPVGLPGVG 272

Qy   169 Q---------------SPRGGAVPRKPVSEMEMERDRGAAHSLEPGKENLPGDP--TSN 210
                           : | || | | | :|| | |
Db   273 KPGVTGFPGPGPQGPLGKPGPGEPGPGLPGVGVKGPGQDGIPGQPGFPGG 332

Qy   211 ATSRGAEGPLPPPSMPAVA-----GAAGGLALLLLGVAGAGGAMCWRRRR------------A 256
          | || ||| || | | | || |||| ||| ||| | | |
Db   333 KGEBQGLPGLPGLPGVGKPGFPGPKGD-----RGIGGVPGALGPRGEKGPVGAPGMGG 387

Qy   257 KPSESRHPG-PGSFGRGGSLGL------------GGGGGMGPR------EAEPGELGIALRGG--GAAD 304
          |||||| | ||| ||| | | ||||| || |||||||||| || ||||
Db   388 PPGEPGLPGIPGPMGGPPGAIGFPGPKGEGGGIVGPGPPGPGPKGKGEPGLQGFPGKPGFLGEVG 447

Qy   305 PP----FCPHYEKVSGDYGH--------------PVYIVQDGPP----------QSPPNI 336
          || | ||| | | | ||| | ||||
Db   448 PPGIRGLPGPIGPKGEAGHKGLPGLPGVPGLLGGPKGEPGIPGDGQLQGPPGI 499

RESULT 42
S23298
collagen alpha 1(VIII) chain - chicken
C;Species: Gallus gallus (chicken)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999
C;Accession: S23298
R;Ninomiya, Y.; Castagnola, P.; Gerecke, D.; Gordon, M.K.; Jacenko, O.; LuValle, P.; McQ
maguchi, N.; Olsen, B.R.
in Extracellular Matrix Genes, Sandell L.J. and Boyd C.D., eds., pp.79-114, Academic Pre
A;Title: The molecular biology of collagens with short triple-helical domains.
A;Reference number: S22243
A;Accession: S23298
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-744 <NIN>
C;Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology
F;617-743/Domain: complement C1q carboxyl-terminal homology <C1Q>

Query Match   5.8%; Score 141; DB 1; Length 744;
Best Local Similarity   23.6%; Pred. No. 0.022;
Matches   83; Conservative   29; Mismatches 108; Indels 132; Gaps 19;

Qy   67 PPGPHSSPNYEFYKLYLVG----------------GAGQ-RRCEAPPAPNLLLTCDRPDLDLRFT 114

RESULT 43
T30165
hypothetical protein F26B1.4 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C;Accession: T30165
R;Lattelle, P.; Wamsley, P.; Kramer, J.
submitted to the EMBL Data Library, November 1996
A;Description: The sequence of C. elegans cosmid F26B1.
A;Reference number: Z20748
A;Accession: T30165
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-305 <LAT>
A;Cross-references: EMBL:U80444; PIDN:AAB37788.1; GSPDB:GN000019; CESP:F26B1.4
A;Experimental source: strain Bristol N2; clone F26B1
C;Genetics:
A;Gene: CESP:F26B1.4
A;Map position: 1
A;Introns: 51/3, 77/1
C;Superfamily: unassigned collagens

Query Match   5.7%; Score 140.5; DB 2; Length 305;
Best Local Similarity   25.8%; Pred. No. 0.009;
Matches   69; Conservative   22; Mismatches 87; Indels 89; Gaps 15;

Qy   114 TIKFQEYSPNLWGHEFRSHHDYYIIATSDG----------TREGLESLQGGVCLTRGMKVL 164
          | |||:|||:|| ||| || ||||||:| ||| ||| | |||| |||||
Db   47 TQHFQETSEKIW----------DDLVFVATGE-TDRHKRHAYLTRAPVVPVTGG--------- 89

Qy   165 LRVGQSPRGGAVPRKPVSEMEMERDRGA----------AHSL---EPGKENLPG------- 205
          | |||| |||| || ||||||||||||| | | ||||| |||
Db   90 FQTGRTANRGCNCRGPPGPPGPGNGEDGCNRFYCNLSHNLPQLPDDGEPGVPGSSGGDDIS 149

Qy   206 --------DP----TSNATSRGAEGPLPPPSMPAVAGAAGGLALLLLGVAGAGGAMCWRRRAK 257
                  || || |||||||| |||||||||| ||||| |||||||||||| | |
Db   150 VLRHDEPGRCAQCPAGPMGPGPPGPGPPGPGVGITGAPG-----LDGIPGRHGTPGY----- 198

Qy   258 PSESRHPG-GSFGRGGSLGLGGGGGGM----------GPREABPGELGIALRGGGAADPPF 307
          |||||| | ||| ||| | || | ||||||||||||||| | ||
Db   199 PGEPGLPGPVGDPGPDGKPGNPGNGVRVRRIPGPPGPRGAB-GDSGV--QGG----PG 250

Qy   308 CPHYEKVSGDYGHPVYIVQDGPPQSPP 334
          | ||| |||| | | | ||| ||
Db   251 SPGLAGAPGDSGAP-----GPQGPP 270

RESULT 44
T22482

Qy   67 PPGPHSSPNYEFYKLYLVG----------------GAGQ-RRCEAPPAPNLLLTCDRPDLDLRFT 114

hypothetical protein F52B11.4 - Caenorhabditis elegans
A;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999
A;Accession: T22482
R;Matthews, L.
submitted to the EMBL Data Library, November 1996
A;Reference number: Z19569
A;Accession: T22482
A;Note: the sequence from Fig. 1
A;Molecule type: DNA
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Residues: 1-304 <WIL>
A;Cross-references: EMBL:Z82268; PIDN:CAB05195.1; GSPDB:GN00022; CBSP:F52B11.4
C;Experimental source: clone F52B11
C;Genetics:
A;Gene: CBSP:F52B11.4
A;Map position: 4
A;Introns: 27/3
C;Superfamily: unassigned collagens

Query Match   5.7%;   Score 140; DB 2; Length 304;
Best Local Similarity   26.8%; Pred. No. 0.0097;
Matches   59; Conservative 16; Mismatches 87; Indels 58; Gaps 10;

Qy   142 DGTRRGLBSLQGGVCL----------------TRGMKVLLRVQGSPRGGAV-----------P 177
Db   87 EGVDQGSEGAQGGSCSGCCLPGAAGPAGTPG-----KPGRPGRPGAAGLPGNPGRPPAQP 141

Qy   178 RKPVSEMPM------BRDRGAAHSLEP-GKENLPGDPTSNATSRGAEGPLPP------PSMP 226
Db   142 CEPITPPPCKPCPGPAGPAPGAPGPQGDPAGPA-PGAP-GQGSGAGPAGPGPAGPGKGASGAGPKGASGAGPAGNP 200

Qy   227 AVAGAAG--GIALLLLGIVAGAGDAMCWRRRRAKPSESRHP---------------GPGSPGRGGSLG 276
Db   201 GQAGAPGGPGADAGSESIPGAPG--------QAGPQGPPGPAGSPGPAGPGPGQGPQGPAAPGGKG 253

Qy   277 LGGGGGMGPREAEPGELGIALRGGGAAADPFCPHYEKVSG 316
Db   254 PSGAPGGPGADGNPGAPGQPGQGAGGAGEKGICPKYCAIDG 293

RESULT 45
CGHU1D
collagen alpha 1(X) chain precursor - human
N;Alternate names: procollagen alpha 1(X) chain
A;Species: Homo sapiens (man)
submitted to the EMBL Data Library, March 1992
A;Reference number: S30086
A;Accession: S30086
A;Molecule type: DNA
A;Residues: 1-680 <APT>
A;Accession: S26396; S30086; S15826; S18249; A43901; S15870; S21856
C;Date: 22-Nov-1993 #sequence revision 03-Nov-1995 #text change 22-Jun-1999
R;Reichenberger, E.; Beier, F.; LuValle, P.; Olsen, B.R.; von der Mark, K.; Bertling, W.
FEBS Lett. 311, 305-310, 1992
A;Title: Genomic organization and full-length cDNA sequence of human collagen X.
A;Reference number: S26396; MUID:93012005; PMID:1397333
A;Accession: S26396
A;Molecule type: DNA
A;Residues: 1-680 <REI>
A;Cross-references: EMBL:X68952; EMBL:X72578; EMBL:X72579; EMBL:X72580; GB:S47714; GB:S4
R;Apte, S.
submitted to the EMBL Data Library, March 1992
A;Reference number: S30085
A;Accession: S30085
A;Molecule type: DNA
A;Residues: 52-680 <AP2>
R;Apte, S.; Mattei, M.G.; Olsen, B.R.
FEBS Lett. 282, 393-396, 1991
A;Title: Cloning of human alpha-1(X) collagen DNA and localization of the COL10A1 gene t
A;Reference number: S15826; MUID:91243838; PMID:2037056
A;Accession: S15826
A;Molecule type: DNA
A;Residues: 561-647,'G',649-666 <AP2>
A;Cross-references: EMBL:X58879; NID:g30013; PIDN:CAA41686.1; PID:g30014
R;Thomas, J.T.; Cresswell, C.J.; Rash, B.; Nicolai, H.; Jones, T.; Solomon, E.; Grant, M
Biochem. J. 280, 617-623, 1991

A;Title: The human collagen X gene. Complete primary translated sequence and chromosomal
A;Reference number: S18249; MUID:92109659; PMID:1764025
A;Accession: S18249
A;Molecule type: DNA
A;Residues: 1-26, 'T', 28-680 <THO>
A;Cross-references: EMBL:X60382; NID:g30094; PIDN:CAA42933.1; PID:g30095
A;Note: the sequence from Fig. 2 in having 532-Ala
R;Reichenberger, E.; Aigner, T.; von der Mark, K.; Stoss, H.; Bertling, W.
Dev. Biol. 148, 562-572, 1991
A;Title: In situ hybridization studies on the expression of type X collagen in fetal hu
A;Reference number: A43901; MUID:92077285; PMID:1743401
A;Accession: A43901
A;Molecule type: mRNA
A;Residues: 547-656 <RE2>
A;Cross-references: GB:M74050; GB:D57494; NID:g339884; PIDN:AAA61221.1; PID:g553796
A;Note: sequence extracted from NCBI backbone (NCBIN:69012, NCBIP:69014)
R;Wallis, G.A.; Rash, B.; Sweetman, W.A.; Thomas, J.T.; Super, M.; Evans, G.; Grant, M.
Am. J. Hum. Genet. 54, 169-178, 1994
A;Title: Amino acid substitutions of conserved residues in the carboxyl-terminal domain
pe Schmid.
A;Reference number: I51870; MUID:94136476; PMID:8304336
A;Accession: I51870
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 520-597,'D',599-680 <WAL>
A;Cross-references: GB:S69531; NID:g545180; PIDN:AAC60615.1; PID:g545181
A;Note: mutant sequence from patient with metaphyseal chondrodysplasia type Schmid
A;Comment: a second mutant sequence with 614-Pro is also described
A;Comment: Prolines and lysines at the third position of the tripeptide repeating unit
ed and subsequently O-glycosylated.
C;Genetics:
A;Gene: GDB:COL10A1
A;Cross-references: GDB:128635; OMIM:120110
A;Map position: 6q21-6q22
A;Introns: 52/1
A;Complex: type X collagen may be a homotrimer
C;Function:
A;Description: structural component of extracellular fibrous polymer specifically and ti
be important for skeletogenesis
C;Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology
C;Keywords: coiled coil; extracellular matrix; glycoprotein; homotrimer; hydroxylysine;
F;1-18/Domain: signal sequence #status predicted <SIG>
F;19-680/Product: collagen alpha 1(X) chain #status predicted <MAT>
F;19-56/Domain: amino-terminal nonhelical #status predicted <NC2>
F;57-519/Region: interrupted helical
F;520-680/Domain: amino-terminal nonhelical #status predicted <NC1>
F;553-679/Domain: complement C1q carboxyl-terminal homology <C1Q>
F;617/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match   5.7%;   Score 140; DB 1; Length 680;
Best Local Similarity   29.1%; Pred. No. 0.024;
Matches   57; Conservative 13; Mismatches 84; Indels 42; Gaps   9;

Qy   171 PRG--------GAVPRKP-------------VSEMPMERDRGAAHSLEPGKKENLPG----------DPTSNA 211
Db   70 PRGHPGPSGPDGKPGYGSPGLGGEPGLPGPGPSAVGKPGVPGLPGKPGERGPYGPKGDV 129

Qy   212 TSRGAEGPLPPSMPAVAGAAGGLALLLLGIVAGAGDAMCWRRRRAKPSESRHPG-PGSFG 270
Db   130 GPAGLPGPRGPPGPGIPGPAG---ISVPGKPGQGPTGAPGPRGPPGPGEKGAPGVPGGMNG 186

Qy   271 RGGSLGLGGGGGMGPREAEPGELGIALRGG--GAADPP-------FCPHYEKVSGDYGH 320
Db   187 QKGEMGYGAPG------RPGERGLPGPQGPTGPGSPGPGVGKRGENGVPGQPGIKGDRGF 239

Qy   321 PVYIVQDGP--PQSPP 334
Db   240 PGEMGPIGPPGPQGPP 255

RESULT 46
B44984

collagen - nematode (Haemonchus contortus) (fragment)
C;Species: Haemonchus contortus
C;Date: 28-Apr-1993 #sequence_revision 28-Apr-1993 #text_change 07-May-1999
C;Accession: B44984
R;Shamansky, L.M.; Pratt, D.; Boisvenue, R.J.; Cox, G.N.
Mol. Biochem. Parasitol. 37, 73-86, 1989
A;Title: Cuticle collagen genes of Haemonchus contortus and Caenorhabditis elegans are h
A;Reference number: A44984; MUID:90136718; PMID:2615789
A;Accession: B44984
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-210 <SHA>
A;Cross-references: GB:J04671; GB:J04670
C;Superfamily: unassigned collagens

Query Match   5.7%;   Score 139.5;   DB 2;   Length 210;
Best Local Similarity   31.1%;   Pred. No. 0.007;
Matches 50;   Conservative 5;   Mismatches 75;   Indels 31;   Gaps 5;

Qy   176 VPRKPVSEMPERDRGAAHSLEPGKENLPGDE---TSNATSRGAEGPLPPPSMPAVAGAAG 233
Db   51 IPIPPCKPCPQGRPGPGPIGPGPEPGTPGNPGAPGNDAPDPGPPGPPGPPGPKAGAP- 109

Qy   234 GIALLLLGVAGAGAGAMCWRRRRAKPSESRHPG---PGSFGRGGSLGLGGGGMGPREAE 289
Db   110 --GAAGPGANA--------PSEPLVPGPGPPGPTPGQGPGPPGPNGAPGHGPAGPA 155

Qy   290 PGELGIALRGGGAAADPPFCPHYEKVSGDYGHPVYIVQDGPP 330
Db   156 PGEKGPRCGDGHPGAP-----------GNAGHGQPQPGPP 186

RESULT 47
JE0291
FB19 protein - human
C;Species: Homo sapiens (man)
C;Date: 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change 21-Jul-2000
C;Accession: JE0291
R;Totaro, A.; Grifa, A.; Carella, M.; Rommens, J.M.; Valentino, M.A.; Roetto, A.; Zeland
Biochem. Biophys. Res. Commun. 250, 555-557, 1998
A;Title: Cloning of a new gene (FB19) within HLA class I region.
A;Reference number: JE0291; MUID:99003493; PMID:9784381
A;Accession: JE0291
A;Molecule type: mRNA
A;Residues: 1-940 <TOT>
C;Genetics:
C;Cross-references: GB:Y13247; NID:g2117158; PIDN:CAA73697.1; PID:g2117159
A;Gene: FB19
A;Map position: 6p21.3

Query Match   23.2%;   Score 139;   DB 2;   Length 940;
Best Local Similarity   5.7%;   Pred. No. 0.04;
Matches 85;   Conservative 15;   Mismatches 111;   Indels 156;   Gaps 18;

Qy   2 GPPHSGPGGVRVGALLLGVLLGVLGVSG-----LSLEPVYWNSANKRFQAEG---- 46
Db   570 GPQGPGGGGGINVQE-ILTSIMGSPNSHPSEELLKQPDYSDKIKQMLVPHGLLGPGPIANG 628

Qy   47 --------GYVLYPQIGDRLDLLCPRARPGHSSPNYEFYKLYLVGGAGGRRCEAPP 96
Db   629 FPPGGPGGPKGMQHFPP---------GPGGPMPGPHGGP---------GGPVGPRLLGPP 670

Qy   97 APNLLLTCDRPLDLRFTIKFQEYSPNLWGHEFRSHHDYYIIATSDGTREGLESLQGGGVC 156
Db   671 PP------------ 672

Qy   157 LTRGMKVLLRVGQSPRGGAV-----PRKPVSEMPERDRGAAHSLEPGKENLPGDPTSNAT 212
Db   673 PRGDPFWDGPGPDEMRGPMRGPGPG-----PG---PGPYHRGRG 708

Qy   213 SRGAEGPLPPPSMPAVAGAAGGIA--LLLLGVAGAGAGAMCWRRRRAKPSESRHPGGSFG 270
Db   709 GRGGNEPPPP--PFFRGARGRGRSGGGGPPNGRGGPGGGMV-----GGGGHRPHEGPGG-G 760

Qy   271 RGGSLG-------LGGGGGMG----PREAEPGELGIALRGGGAADPPFCPHYEKVSGDYGHPV 322
Db   761 MGNSGGHRPHEGPGGGGMGSGHRPHEGPGGGSSMG-----GGGGHRPHEGPG-GGISGGSSGHRP 815

Qy   323 YIVQDGP 329
Db   816 H---EGP 819

RESULT 48
S18251
collagen alpha 1(XI) chain - bovine (fragments)
C;Species: Bos primigenius taurus (cattle)
C;Date: 22-Nov-1993 #sequence_revision 23-Mar-1995 #text_change 02-Jul-1998
A;Accession: S18251; C46662; A56978; S65864; D46662; E46662; F46662; G46662; H46662; I4
R;Brown, K.E.; Lawrence, R.; Sonenshein, G.E.
J. Biol. Chem. 266, 23268-23273, 1991
A;Title: Concerted modulation of alpha-1(XI) and alpha-2(V) collagen mRNAs in bovine va
A;Reference number: S18251; C46662; A56978; S65864; D46662; MUID:92078200; PMID:1744123
A;Accession: S18251
A;Molecule type: mRNA
A;Residues: 1-911 <BRO>
R;Mayne, R.; Brewton, R.G.; Mayne, P.M.; Baker, J.R.
J. Biol. Chem. 268, 9381-9386, 1993
A;Title: Isolation and characterization of the chains of type V/type XI collagen presen
A;Reference number: A46662; MUID:93252802; PMID:8486632
A;Accession: C46662
A;Molecule type: protein
A;Residues: 347-354,356-363,586-600,912-924,925-961,962-998,999-1024 <MAY>
A;Experimental source: vitreous humor
A;Note: sequence modified after extraction from NCBI backbone
R;Wu, J.J.; Eyre, D.R.
J. Biol. Chem. 270, 18865-18870, 1995
A;Title: Structural analysis of cross-linking domains in cartilage type XI collagen. Ir
A;Reference number: A56978; MUID:95370194; PMID:7642541
A;Accession: A56978
A;Molecule type: protein
A;Status: preliminary
A;Residues: 205-229 <WUA>
R;Niyibizi, C.; Eyre, D.R.
Eur. J. Biochem. 224, 943-950, 1994
A;Title: Structural characteristics of cross-linking sites in type V collagen of bone.
A;Reference number: S48210; MUID:95010086; PMID:7925418
A;Accession: S65864
A;Molecule type: protein
A;Residues: 273-298 <NIY>
C;Superfamily: collagen alpha 1(V) chain; fibrillar collagen carboxyl-terminal homology
C;Keywords: hydroxyproline
F;211,223/Modified site: 4-hydroxyproline (pro) #status experimental

Query Match   24.3%;   Score 139;   DB 2;   Length 1024;
Best Local Similarity   5.7%;   Pred. No. 0.044;
Matches 90;   Conservative 19;   Mismatches 118;   Indels 144;   Gaps 19;

Qy   2 GPPHSGPGGVRVGALLLGVLLGVLGVSGLSLEPVYWNSANKRFQAEGGGYVLYPQIGDRLDLL 61
Db   728 GPP---GPGGV-VGPGQPTGBTG-----------------------------PIGER---- 751

Qy   62 CPRARPPGPHSSPNYEFYKLYLVGGAGGRRCEAPPAPNLLLTCDRPLDLRFTIKFQEYS 121
Db   752 -GHPGPPGPGBQG-----LPGAAGKESAGKGPDGPQGVSGKDGP--------- 789

Qy   122 PNLWGHEFRSHHDYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKPV 181
Db   790 AGLRGF--------PGER--GLPGAQAGPL--- 820

Qy   182 SEMPMERDRGAAHSL------EPGKENLPGDPTSNAISRGAEGPLPPPSMPAVAGAAGG 234
Db   821 GPVGSPGERGSAGTAGPIGLPGRGPQGPPG-PAGEKGPAPGEKGPQGPAGRDGVQGPVG- 878

Qy   235 LALLLLGVAGAGAGAMCWRRRRAKPSESRHPG--------------PGSFGRG----- 272
Db   879 ...

```
Db      879  ---------LPGPAGPAGS--------------PGEDGDKGEIGEPGQKGSKGDKEGQQGMPGQGDEAG  925

Qy      273  -----GSLGLGGGGGMGPREAEPGELGIALRGGGAAADPPFCPHYEKVSGDYGHPVYIVQD  327
                  |:|:||                              :|:|
Db      926  PPGPGPIGLQGLPG--PPGGENGDVGPM---GPPGPP-GPGIPGPVGEQGLPGAAGQD  979

Qy      328  GP--PQSPNI  336
Db      980  GPPGPLGPPGL  990
```

```
RESULT 49
A24450
collagen alpha 2(VIII) chain - bovine (fragment)
N;Alternate names: 50K-A collagen
C;Species: Bos primigenius taurus (cattle)
C;Date: 31-Mar-1988 #sequence_revision 05-Apr-1995 #text_change 03-May-1996
R;Mann, K.; Jander, R.; Korsching, E.; Kuehn, K.; Rauterberg, J.
FEBS Lett. 273, 168-172, 1990
A;Title: The primary structure of a triple-helical domain of collagen type VIII from bov
A;Reference number: S12898; MUID:91032171; PMID:2226849
A;Accession: S12898
A;Status: preliminary
A;Molecule type: Protein
A;Residues: 1-469 <MAN>
R;Kapoor, R.; Bornstein, P.; Sage, E.H.
Biochemistry 25, 3930-3937, 1986
A;Title: Type VIII collagen from bovine Descemet's membrane: structural characterization
A;Reference number: A90507; MUID:86296625; PMID:3527259
A;Accession: A24450
A;Molecule type: protein
A;Residues: 1-24 <KAP>
C;Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology

Query Match     5.7%;  Score 138.5;  DB 2;  Length 469;
Best Local Similarity  25.3%;  Pred. No. 0.02;
Matches  69;  Conservative  17;  Mismatches  72;  Indels  115;  Gaps  14;
```

```
Qy      147  GLESLGQ-------GVCLTRGMKVLLRVGQSPRGGAVP----------RKPVSBMPMERDRGAA  193
                 |||                     |:|                   :|
Db      216  GLPGPQGPPGAXGEPGTRGPPGLI------GPTGYGMPGLPGPXGDRGPAGPVGPLLGDRG--  269

Qy      194  HSLEPGKENLPGD------------PTSNAT----------SRGAGGPLPPSMPAVAGAAG  233
                 ||::|| :|                            ||:| |||:
Db      270  ---EPGEDGEPGEQGPGLGGPPGLPGSAGLPGRRGVPGPXGETGPIGPPGPVGPGIRGDQG  326

Qy      234  --GLALLLLGVAGAGAGGAMCWRRRAKPSESRHPGP-------------------GSFGRG  272
                |:| |:                                               :||| |
Db      327  PSGLA-------GKPGLPG-------ERGLPGAHGPPGPTGPXGEPGFTGRPGPGAGALGQX  376

Qy      273  GSLGL-------------------------------------EAEPGELGIALRGG  300
                 ||:|                                        :| |
Db      377  GDLGLPGQPGLRGPSGGIPGLQGPAGPIGPGQGPLXGEPGLPGPPGPGEGKVGEPGVA----  432

Qy      301  GAADPPFCPHYEKVSGDYGHPVYIVQDGPPGSP  333
                 ||:                    |:|
Db      433  GPTGPPGVPGGSPGLTGPPGPP-----GPGPP  459
```

```
RESULT 50
A41182
collagen alpha 1(II) chain precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 28-May-1992 #sequence_revision 28-May-1992 #text_change 13-Aug-1999
C;Accession: A41182; A44885
R;Metsaaranta, M.; Toman, D.; de Crombrugghe, B.; Vuorio, E.
J. Biol. Chem. 266, 16862-16869, 1991
A;Title: Mouse type II collagen gene. Complete nucleotide sequence, exon structure, and
A;Reference number: A41182; MUID:91358489; PMID:1885613
A;Accession: A41182
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
```

```
A;Residues: 1-1419 <MET>
A;Cross-references: GB;M65161
R;Cheah, K.S.; Lau, E.T.; Au, P.K.; Tam, P.P.
Development 111, 945-953, 1991
A;Title: Expression of the mouse alpha 1(II) collagen gene is not restricted to cartila
A;Reference number: A44885; MUID:91347939; PMID:1879363
A;Accession: A44885
A;Molecule type: DNA
A;Residues: 1-28 <CHE>
A;Cross-references: GB;863190; NID:g2343368; PIDN:AAB19627.1; PID:g2343369
A;Note: sequence extracted from NCBI backbone (NCBIN:63190, NCBIP:63190, NCBIP:63192)
C;Superfamily: collagen alpha 1(II) chain; fibrillar collagen carboxyl-terminal homology
C;Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trim
F;1191-1419/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match     5.7%;  Score 138.5;  DB 2;  Length 1419;
Best Local Similarity  30.1%;  Pred. No. 0.068;
Matches  58;  Conservative  13;  Mismatches  75;  Indels  47;  Gaps  9;
```

```
Qy      173  GGAVPRKPVSBMPMERDRGAAHSLEPGKENL-------PGD--PTSNATSRGABGPLPPPS  224
                 ||                         |:|        :|||||
Db      430  GGAGPIGPPGERGAPGNRGF------PCQDGLAGPKGAPGERGPSGLAGPKGANGDPGRPG  484

Qy      225  MPAVAGAAGGLALLLLGVAGAGAGGAMCWRRRRAKPSESRHPGP----GSFGRGGSLGLGGG  280
                 ||                                          :||| |
Db      485  EPGILPGARG-----LTGRPGDAGPQGKVGPSGAPGEDGRPGPGPQGRARGQPGVMGFPGP  539

Qy      281  GGMGPREAEPGELGIA-----LRG-------GGAADPPFCPHYEKVSGDYGHPVYIVQDGP  329
                 |:|                 :|          ||:|                |:|
Db      540  KGANGEPGKAGEKGLAGAPGLRGLPGKDGETGAAGPPGPSGPRAGERGPGGAP------GP  593

Qy      330  -------PQSPP  334
                      |||
Db      594  SGFQGLPGPGPGP  606
```

```
Search completed: February 11, 2003, 12:05:56
Job time : 46.195 secs
```

OM protein - protein search, using sw model

Copyright (c) 1993 - 2003 Compugen Ltd.

GenCore version 5.1.3

Run on: February 11, 2003, 12:05:10 ; Search time 19.4591 Seconds
(without alignments)
597.393 Million cell updates/sec

Title: US-10-021-121-2

Perfect score: 2450

Sequence: 1 MGPPHSGPGGVRVGALLLG.........TTLLRQRASVEARGQHGPL 455

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 140259 seqs, 25548876 residues

Total number of hits satisfying chosen parameters: 140259

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : Published Applications AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2450	100.0	455	12	US-10-021-121-2	Sequence 2, Appli
2	1841	75.1	340	12	US-10-021-121-4	Sequence 4, Appli
3	1837	75.0	340	9	US-09-925-297-510	Sequence 3, Appli
4	623	25.4	346	12	US-10-021-121-9	Sequence 9, Appli
5	620.5	25.4	333	10	US-09-754-105-2	Sequence 2, Appli
6	620.5	25.3	333	10	US-09-978-333-2	Sequence 2, Appli
7	620.5	25.3	333	12	US-10-021-121-10	Sequence 10, Appl
8	613.5	25.0	333	12	US-10-138-787-4	Sequence 4, Appli
9	599.5	24.5	345	9	US-09-978-333-10	Sequence 5, Appli
10	599.5	20.0	333	9	US-10-138-787-13	Sequence 13, Appli
11	489	20.0	89	10	US-09-862-179A-17	Sequence 17, Appli
11	489	20.0	89	10	US-09-864-761-48262	Sequence 13, Appli
12	284.5	11.6	92	10	US-09-864-761-48262	Sequence 48262, A
13	196.5	10.0	136	10	US-09-864-761-48257	Sequence 48257, A
14	193.5	7.9	106	10	US-09-925-297-639	Sequence 639, App
15	192	7.8	82	9	US-10-138-787-11	Sequence 11, Appl
16	192	7.8	82	9	US-09-862-179A-15	Sequence 15, Appl
17	191.5	7.8	82	10	US-09-862-179A-12	Sequence 12, Appl
18	191.5	7.8	82	10	US-09-862-179A-16	Sequence 16, Appl
19	179	7.3	238	10	US-09-904-954-2	Sequence 2, Appli
20	176	7.2	209	9	US-09-921-984-2	Sequence 2, Appli
21	174.5	7.1	233	9	US-10-138-787-7	Sequence 7, Appli
22	172.5	7.0	218	10	US-09-925-297-510	Sequence 510, App
23	169.5	6.9	201	10	US-09-904-954-4	Sequence 4, Appli
24	169	6.9	201	9	US-10-138-787-8	Sequence 8, Appli
25	168.5	6.9	209	9	US-10-138-787-6	Sequence 6, Appli
26	167.5	6.8	228	8	US-08-578-684-4	Sequence 4, Appli
27	166	6.8	205	9	US-10-138-787-10	Sequence 10, Appl
28	164.5	6.7	204	9	US-10-028-072-288	Sequence 288, App
29	164.5	6.7	204	9	US-10-121-049-288	Sequence 288, Appl
30	164.5	6.7	204	9	US-10-123-904-288	Sequence 288, App
31	164.5	6.7	204	9	US-10-140-470-288	Sequence 288, App
32	164.5	6.7	204	9	US-10-175-746-288	Sequence 288, App
33	164.5	6.7	204	9	US-10-176-918-288	Sequence 288, App
34	164.5	6.7	204	9	US-10-176-921-288	Sequence 288, App
35	160.5	6.6	204	9	US-09-925-301-1225	Sequence 1225, Ap
36	145	5.9	228	10	US-09-862-179A-2	Sequence 2, Appli
37	145	5.9	27	10	US-10-001-887-108	Sequence 108, Appl
38	141	5.8	28	10	US-09-823-240-11	Sequence 2, Appli
39	140.5	5.7	638	10	US-09-823-240-10	Sequence 108, Appl
40	130.5	5.3	380	10	US-09-919-497-56	Sequence 56, Appl
41	130.5	5.3	674	10	US-09-925-299-979	Sequence 979, Appl
42	129	5.3	595	10	US-09-854-133-187	Sequence 187, App
43	129	5.3	595	10	US-09-738-973-187	Sequence 187, App
44	129	5.3	1806	10	US-09-919-497-56	Sequence 10, Appli
45	127	5.2	430	10	US-09-925-240-9	Sequence 56, Appl
46	127	5.2	684	10	US-09-823-240-9	Sequence 9, Appli
47	126.5	5.2	403	10	US-09-978-295A-614	Sequence 2, Appli
48	126.5	5.2	520	9	US-09-978-697-614	Sequence 614, App
49	126.5	5.2	520	9	US-09-925-302-689	Sequence 614, App
50	126.5	5.2	520	9	US-09-978-192A-614	Sequence 689, App

ALIGNMENTS

RESULT 1
US-10-021-121-2
Sequence 2, Application US/10021121
Patent No. US20020142444A1
GENERAL INFORMATION:
APPLICANT: Caras, Ingrid W
TITLE OF INVENTION: A2-1 Neurotrophic Factor
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 1 DNA Way
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/021,121
FILING DATE: 06-Dec-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/635,130
FILING DATE: 19-Mar-1996
ATTORNEY/AGENT INFORMATION:
NAME: Torchia, PhD., Timothy E.
REGISTRATION NUMBER: 36,700
REFERENCE/DOCKET NUMBER: P1001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-8674
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 455 amino acids

;
;
;   TYPE: Amino Acid
;   TOPOLOGY: Linear
;   SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-10-021-121-2

Query Match   100.0%; Score 2450; DB 12; Length 455;
Best Local Similarity   100.0%; Pred. No. 2.5e-171;
  Matches 455; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 MGPPHSGPGGVRVGALLLLGVLLGVLVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDL 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MGPPHSGPGGVRVGALLLLGVLLGVLVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDL 60

Qy   61 LCPRARPPGPHSSPNYEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDLLDLRFTIKFQEY 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LCPRARPPGPHSSPNYEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDLLDLRFTIKFQEY 120

Qy 121 SPNLWGHEFRSHHDYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKP 180
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 SPNLWGHEFRSHHDYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKP 180

Qy 181 VSEMPMERDRGAAHSLEPGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAGGLALLLL 240
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 VSEMPMERDRGAAHSLEPGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAGGLALLLL 240

Qy 241 GVAGAGGAMCWRRRRAKPSESRHPGPGSFGRGGSLGLGGGGGMGPREAEPGELGIALRGG 300
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 GVAGAGGAMCWRRRRAKPSESRHPGPGSFGRGGSLGLGGGGGMGPREAEPGELGIALRGG 300

Qy 301 GAADPPFCPHYEKVSGDYGHPVYIVQDGPPQSPPNIYYTSISVLEWPILHTIQLFPMRSK 360
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 301 GAADPPFCPHYEKVSGDYGHPVYIVQDGPPQSPPNIYYTSISVLEWPILHTIQLFPMRSK 360

Qy 361 CSRVTTLFPVQVITTSTCRMTSFSFTTLNPSMQACRAQMGEFRIRWCFWGDRILGTALF 420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 361 CSRVTTLFPVQVITTSTCRMTSFSFTTLNPSMQACRAQMGEFRIRWCFWGDRILGTALF 420

Qy 421 VLVLILLIGRLNMHQTTLLRQRASVEABAGQHGPL 455
        |||||||||||||||||||||||||||||||||||
Db 421 VLVLILLIGRLNMHQTTLLRQRASVEABAGQHGPL 455

RESULT 2
US-10-021-121-4
; Sequence 4, Application US/10021121
; Patent No. US2002014244A1
; GENERAL INFORMATION:
;   APPLICANT: Caras, Ingrid W
;   TITLE OF INVENTION: A2-1 Neurotrophic Factor
;   NUMBER OF SEQUENCES: 10
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Genentech, Inc.
;   STREET: 1 DNA Way
;   CITY: South San Francisco
;   STATE: California
;   COUNTRY: USA
;   ZIP: 94080
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: WinPatin (Genentech)
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/10/021,121
;   FILING DATE: 06-Dec-2001
;   CLASSIFICATION: <Unknown>
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US/08/635,130
;   FILING DATE: 19-Mar-1996
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Torchia, PhD., Timothy E.
;   REGISTRATION NUMBER: 36,700
;   REFERENCE/DOCKET NUMBER: P1001

;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 650/225-8674
;   TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 4:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 340 amino acids
;   TYPE: Amino Acid
;   TOPOLOGY: Linear
;   SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-10-021-121-4

Query Match   75.1%; Score 1841; DB 12; Length 340;
Best Local Similarity   100.0%; Pred. No. 4.3e-127;
  Matches 338; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 MGPPHSGPGGVRVGALLLLGVLLGVLVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDL 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MGPPHSGPGGVRVGALLLLGVLLGVLVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDL 60

Qy   61 LCPRARPPGPHSSPNYEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDLLDLRFTIKFQEY 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LCPRARPPGPHSSPNYEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDLLDLRFTIKFQEY 120

Qy 121 SPNLWGHEFRSHHDYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKP 180
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 SPNLWGHEFRSHHDYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKP 180

Qy 181 VSEMPMERDRGAAHSLEPGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAGGLALLLL 240
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 VSEMPMERDRGAAHSLEPGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAGGLALLLL 240

Qy 241 GVAGAGGAMCWRRRRAKPSESRHPGPGSFGRGGSLGLGGGGGMGPREAEPGELGIALRGG 300
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 GVAGAGGAMCWRRRRAKPSESRHPGPGSFGRGGSLGLGGGGGMGPREAEPGELGIALRGG 300

Qy 301 GAADPPFCPHYEKVSGDYGHPVYIVQDGPPQSPPNIYY 338
        ||||||||||||||||||||||||||||||||||||||
Db 301 GAADPPFCPHYEKVSGDYGHPVYIVQDGPPQSPPNIYY 338

RESULT 3
US-10-138-787-3
; Sequence 3, Application US/10138787
; Patent No. US2002017296A1
; GENERAL INFORMATION:
;   APPLICANT: Mbanalu, Geraldine
;   APPLICANT: Pawson, Tony
;   APPLICANT: Holland, Sacha
;   APPLICANT: Mbanalu, Geraldine
;   TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED
;   TITLE OF INVENTION: BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR
;   TITLE OF INVENTION: TYROSINE KINASES
;   FILE REFERENCE: 11757.23USWO
;   CURRENT APPLICATION NUMBER: US/10/138,787
;   CURRENT FILING DATE: 2002-05-03
;   PRIOR APPLICATION NUMBER: US/09/214,631
;   PRIOR FILING DATE: 1999-03-12
;   PRIOR APPLICATION NUMBER: PCT/CA97/00473
;   PRIOR FILING DATE: 1997-07-04
;   PRIOR APPLICATION NUMBER: US/10/138,787
;   PRIOR FILING DATE: 1997-07-04
;   PRIOR APPLICATION NUMBER: 60/021,272
;   PRIOR FILING DATE: 1996-07-05
;   NUMBER OF SEQ ID NOS: 13
;   SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 340
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-138-787-3

Query Match   75.0%; Score 1837; DB 9; Length 340;
Best Local Similarity   99.7%; Pred. No. 8.4e-127;
  Matches 337; Conservative   1; Mismatches   0; Indels   0; Gaps   0;

Qy   1 MGPPHSGPGGVRVGALLLLGVLIGLVLVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDL 60

```
RESULT 4
US-10-021-121-9
; Sequence 9, Application US/10021121
; Patent No. US20020142444A1
; GENERAL INFORMATION:
;   APPLICANT: Carae, Ingrid W
;   TITLE OF INVENTION: A2-1 Neurotrophic Factor
;   NUMBER OF SEQUENCES: 10
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Genentech, Inc.
;     STREET: 1 DNA Way
;     CITY: South San Francisco
;     STATE: California
;     COUNTRY: USA
;     ZIP: 94080
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: WinPatin (Genentech)
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/10/021,121
;     FILING DATE: 06-Dec-2001
;     CLASSIFICATION: <Unknown>
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US/08/635,130
;     FILING DATE: 19-Mar-1996
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Torchia, PhD., Timothy E.
;     REGISTRATION NUMBER: 36,700
;     REFERENCE/DOCKET NUMBER: P1001
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 650/225-8674
;     TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 9:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 346 amino acids
;     TYPE: Amino Acid
;     TOPOLOGY: Linear
;   SEQUENCE DESCRIPTION: SEQ ID NO: 9:

US-10-021-121-9
Query Match             25.4%;   Score 623;   DB 12;   Length 346;
Best Local Similarity   39.2%;   Pred. No. 2.5e-38;
Matches  143;   Conservative  48;   Mismatches 116;   Indels  58;   Gaps  9;

Qy      8 PGGVRVGALLLLGVLLGVLSGL------SLEPVYWNSANKRFQAEGGYVLYPQIGDRLDLL   61
            ||: :|::   |::|: :||:||      :|:|||: :|::||  ||||:::|||||||
Db      4 PGQRWLGKWLVAMVVWALCRLATPLAKNLEPVSNSSLNPKFLSGKGLVIYPKIGDKLDII   63
```

```
RESULT 5
US-09-754-105-2
; Sequence 2, Application US/09754105
; Patent No. US20010009768A1
; GENERAL INFORMATION:
;   APPLICANT: Cerretti, Douglas
;   APPLICANT: Reddy, Prahnitha
;   TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING CYTOKINE DESIGNATED LERK-5
;   FILE REFERENCE: 28232
;   CURRENT APPLICATION NUMBER: US/09/754,105
;   CURRENT FILING DATE: 2001-01-03
;   PRIOR APPLICATION NUMBER: 09/329,531
;   PRIOR FILING DATE: 1999-06-10
;   NUMBER OF SEQ ID NOS: 3
;   SOFTWARE: Patentin version 3.0
; SEQ ID NO 2
;   LENGTH: 333
;   TYPE: PRT
;   ORGANISM: homo sapiens
US-09-754-105-2

Query Match             25.3%;   Score 620.5;   DB 10;   Length 333;
Best Local Similarity   40.5%;   Pred. No. 3.7e-38;
Matches  133;   Conservative  52;   Mismatches 130;   Indels  13;   Gaps  5;

Qy     14 GALLLLGVLLGVLSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDLLCPRARPPGPHSS   73
            |:|:    |:| |||:|  |||||||||| |:   | ::  :  |  ||||:||||||||
Db     14 GVLMVLCRTAISKSIVLEPIYWNSNSKFLPGQGLVLYPQIGDRLDIICPKV---DSKTV   70

Qy     74 PNYEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDLDLRFTIKFQEYSPNLWGHEFRSHH  133
            |:||| |||:  | : |: |:|  ::||| ||:|||| ||||||: |||||| ||:|||
Db     14 GQYEYVKYVMDKDQADRCIIKKENTPLLNCAKPDQDIKFIKFQEFSPNLWGLEFQKNK  130

Qy    134 DYYIIATSDGTREGLESLQGGVCLITRGMKVLLRVGQ--SPRGGAVPRKPVSEMPMER-DR  190
            |||:|:|:|  ||:| ||:|::    |::|  |||:|     |:  | :|||||||: |:
Db    131 DYYIISTNGSLEGLDNQBEGGVCQTRAMKILMKVGQDASSAGSTRNKDPTRPELEAGTN  190

Qy    191 GAAHSLEPGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAGGLALLLLLGVLAGAGAMC  250
            |||:||||||||||||||||||||||:||:|||||||||||||| ||:||||:||||||:|
Db    191 GRSSTTSPFVKENPGGSSTDGNSAGHSGNNILGSEVALFAGIASGCIIFIVIIITLVVLL  250

Qy    251 WRRRRAKPSESRHPGPGSFGRGGSLGLGGGGMGPREAEPGELGIALRGGAADPPFCPH  310
            |||||||||||||||||||||||||||| ||:     |     :|:||:|     :::|
Db    251 KYRRHRKHSPQHTTLSLSTLATPKRSGNN------NGSEPSDIIIPLR---TADSVPCPH  303

Qy    311 YEKVSGDYGHPVYIVQDGPPQSPPNIYY  338
            |||||||||||||||||||||||||||||
Db    311 YEKVSGDYGHPVYIVQDGPPQSPPNIYY  338
```

Left column

Db 304 YEKVSGDYGHPVYIVQEMPQSPANIYY 331

RESULT 6
US-09-978-339-2
; Sequence 2, Application US/09978339
; Patent No. US20020103358A1
; GENERAL INFORMATION:
;   APPLICANT: Cerretti, Douglas P.
;   Reddy, Prashitha
;   TITLE OF INVENTION: No. US20020103358A1e1 Cytokine Designated Lerk-5
;   NUMBER OF SEQUENCES: 3
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Immunex Corporation
;   STREET: 51 University Street
;   CITY: Seattle
;   STATE: Washington
;   COUNTRY: US
;   ZIP: 98101
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: Apple Macintosh
;   OPERATING SYSTEM: Apple 7.1
;   SOFTWARE: Microsoft Word, Version 5.1a
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/978,339
;   FILING DATE: 15-Oct-2001
;   CLASSIFICATION: <Unknown>
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/271,948
;   FILING DATE: <Unknown>
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Seese, Kathryn A.
;   REGISTRATION NUMBER: 32,172
;   REFERENCE/DOCKET NUMBER: 2823
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (206) 587-0430
;   TELEFAX: (206) 233-0644
;   TELEX: 756822
; INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 333 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
;   SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-978-339-2

Query Match   25.3%; Score 620.5; DB 10; Length 333;
Best Local Similarity 40.5%; Pred. No. 3.7e-38;
Matches 133; Conservative 52; Mismatches 130; Indels 13; Gaps 5;

Qy   14 GALILLGVLGLVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDLLCPRARPPGPHSS 73
Db   14 GVLMVLCRTAISKSIVLEPIYWNSSNSKFLPGQGLVLYPQIGDKLDIICPKV---DSKTV 70

Qy   74 PNYEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDLDLRFTIKFQEYSPNLWGHEFRSHH 133
Db   71 GQYEYKVYMVDKDQADRCITKKENTPLINCAKPDQDIKFTIKFQEFSPNLWGLEFQKNK 130

Qy 134 DYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQ--SPRGGAVPRKPVSEMMER-DR 190
Db 131 DYYIISTNSGSLEGLDNQEGGVCQTRAMKILMKVGQDASSAGSTRNKDPTRRPELEAGTN 190

Qy 191 GAAHSLEPGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAGGLALLLLGVAGAGGAMC 250
Db 191 GRSSTTSPFVKPNPGSSTDGNSAGHSGNNILGSEVALFAGIIASGCIIFIVIIITLVVLLL 250

Qy 251 WRRRAKPSESRHPGPGSFGRGGSLGLGGCGGMGPREAEPGELGIALRGGGAADPPFCPH 310
Db 251 KYRRHRKHSPQHTTTLSLTLATPKRSGNN----NGSEPSDIIIPLR---TADSVFCPH 303

Qy 311 YEKVSGDYGHPVYIVQDGPPQSPPNIYY 338
Db 304 YEKVSGDYGHPVYIVQEMPQSPANIYY 331

Right column

Db 304 YEKVSGDYGHPVYIVQEMPQSPANIYY 331

RESULT 7
US-10-021-121-10
; Sequence 10, Application US/10021121
; Patent No. US20020142444A1
; GENERAL INFORMATION:
;   APPLICANT: Caras, Ingrid W
;   TITLE OF INVENTION: A2-1 Neurotrophic Factor
;   NUMBER OF SEQUENCES: 10
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Genentech, Inc.
;   STREET: 1 DNA Way
;   CITY: South San Francisco
;   STATE: California
;   COUNTRY: USA
;   ZIP: 94080
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: WinPatin (Genentech)
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/10/021,121
;   FILING DATE: 06-Dec-2001
;   CLASSIFICATION: <Unknown>
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US/08/635,130
;   FILING DATE: 19-Mar-1996
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Torchia, PhD., Timothy E.
;   REGISTRATION NUMBER: 36,700
;   REFERENCE/DOCKET NUMBER: P1001
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 650/225-8674
;   TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 10:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 333 amino acids
;   TYPE: Amino Acid
;   TOPOLOGY: Linear
;   SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-10-021-121-10

Query Match   25.3%; Score 620.5; DB 12; Length 333;
Best Local Similarity 40.5%; Pred. No. 3.7e-38;
Matches 133; Conservative 52; Mismatches 130; Indels 13; Gaps 5;

Qy   14 GALILLGVLGLVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDLLCPRARPPGPHSS 73
Db   14 GVLMVLCRTAISKSIVLEPIYWNSSNSKFLPGQGLVLYPQIGDKLDIICPKV---DSKTV 70

Qy   74 PNYEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDLDLRFTIKFQEYSPNLWGHEFRSHH 133
Db   71 GQYEYKVYMVDKDQADRCITKKENTPLINCAKPDQDIKFTIKFQEFSPNLWGLEFQKNK 130

Qy 134 DYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQ--SPRGGAVPRKPVSEMMER-DR 190
Db 131 DYYIISTNSGSLEGLDNQEGGVCQTRAMKILMKVGQDASSAGSTRNKDPTRRPELEAGTN 190

Qy 191 GAAHSLEPGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAGGLALLLLGVAGAGGAMC 250
Db 191 GRSSTTSPFVKPNPGSSTDGNSAGHSGNNILGSEVALFAGIIASGCIIFIVIIITLVVLLL 250

Qy 251 WRRRAKPSESRHPGPGSFGRGGSLGLGGCGGMGPREAEPGELGIALRGGGAADPPFCPH 310
Db 251 KYRRHRKHSPQHTTTLSLTLATPKRSGNN----NGSEPSDIIIPLR---TADSVFCPH 303

Qy 311 YEKVSGDYGHPVYIVQDGPPQSPPNIYY 338
Db 304 YEKVSGDYGHPVYIVQEMPQSPANIYY 331

RESULT 8
US-10-138-787-4
; Sequence 4, Application US/10138787
; Patent No. US20020172984A1
; GENERAL INFORMATION:
; APPLICANT: Holland, Sacha
; APPLICANT: Mbamalu, Geraldine
; APPLICANT: Pawson, Tony
; TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED
; TITLE OF INVENTION: BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR
; TITLE OF INVENTION: TYROSINE KINASES
; FILE REFERENCE: 11757.23USWO
; CURRENT APPLICATION NUMBER: US/10/138,787
; CURRENT FILING DATE: 2002-05-03
; PRIOR APPLICATION NUMBER: US/09/214,631
; PRIOR FILING DATE: 1999-03-12
; PRIOR APPLICATION NUMBER: PCT/CA97/00473
; PRIOR FILING DATE: 1997-07-04
; PRIOR APPLICATION NUMBER: 60/021,272
; PRIOR FILING DATE: 1996-07-05
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-138-787-4

Query Match   25.0%; Score 613.5; DB 9; Length 333;
Best Local Similarity   40.5%; Pred. No. 1.2e-37;
Matches 133; Conservative 54; Mismatches 128; Indels 13; Gaps   6;

Qy   14 GALLLLGVLGLVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLLDLLCPRARPPGPHSS 73
                ::|: ||||||||| :|| ::|||||||||||||||||||:|||||| :||||||||
Db   14 GVLMVLCRTAISKSIVLEPIYMSSNSKFLPGQGLVLYPQIGDKLDIICPKV---DSKIV 70

Qy   74 PNYEFYKLYLVGGAQGRRCEAPPAPNLLITCDRPDLDLRFTIKFQEYSPNLWGHEFRSHH 133
                |:|:| ||||:|:|: |::|| :|||||||||||||||||||||::|||||:|:
Db   71 GQYEYKVKVMVDKDQADRCTIKENTPLLNCAKPDQDIKFTIKPQEFSPNLWGLEFQKNK 130

Qy   134 DYYIIATSDGTREGCLESLQGGVCLLTRGMKVLLRVGQ-SPRGGAVPRKPVSEMP-MER-DR 190
                ||||||||||:|:|||||||||||||||||||||| :|| ::|||||| ||: ||
Db   131 DYYIISTSNGSLEGLDNQEGGVCQTRAMKILMKVGQDASSAGSTPNKDPTRRPELEAGTN 190

Qy   191 GAAHSLEPGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAGGLALLLLGVAGAGGAMC 250
                : ||||||||||||||||||||||||||||||: ::||
Db   191 GRSSTTSPFVKPNPGSSTDGNSAGHSGNNILGSEVALFAGIASGCIFIVIITLVVLL 250

Qy   251 WRRRAKPSESRHPGPGSFGRGGSLGLGGGGMGPREAEPGELGIALRGGGAADPPFCPH 310
                |||||||||||:|||||:|: ||||||: || ||||
Db   251 KYRRHRKHSPQHTTLSLSTLATPKRSGNN----NGSEPSDIIIPLR---TADSVFCPH 303

Qy   311 YEKVSGDYGHPVYIVQDGPPQSPPNIYY 338
                |||||||||||:|||||||| :||| ||||
Db   304 YEKVSGDYGHPVVIVQEMPPQSPANIYY 331

;

RESULT 9
US-10-138-787-5
; Sequence 5, Application US/10138787
; Patent No. US20020172984A1
; GENERAL INFORMATION:
; APPLICANT: Holland, Sacha
; APPLICANT: Mbamalu, Geraldine
; APPLICANT: Pawson, Tony
; TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED
; TITLE OF INVENTION: BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR
; TITLE OF INVENTION: TYROSINE KINASES
; FILE REFERENCE: 11757.23USWO
; CURRENT APPLICATION NUMBER: US/10/138,787
; CURRENT FILING DATE: 2002-05-03

; PRIOR APPLICATION NUMBER: US/09/214,631
; PRIOR FILING DATE: 1999-03-12
; PRIOR APPLICATION NUMBER: PCT/CA97/00473
; PRIOR FILING DATE: 1997-07-04
; PRIOR APPLICATION NUMBER: 60/021,272
; PRIOR FILING DATE: 1996-07-05
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-138-787-5

Query Match   24.5%; Score 599.5; DB 9; Length 345;
Best Local Similarity   38.4%; Pred. No. 1.3e-36;
Matches 140; Conservative 50; Mismatches 116; Indels 59; Gaps 10;

Qy   8 PGGVRVG--------ALLLLGVLGLVSGL--SLEPVYWNSANKRFQAEGGYVLYPQIGDRLLDLL 61
                :: ::|::||||||||| :|| ::|||||||||||||||||||:|||||
Db   4 PGGRWLCKWLYAVVWALCRLATPLARKNLEPVSWSSSLNPKFLSGKGLVIYPKIGDKLDII 63

Qy   62 CPRARPPGPHSSPNYEFYKLYLVGGAQGRRCEAPPAPNLLITCDRPDLDLRFTIKFQEYS 121
                :||||:||| ||:|::|||| ||:|||:||| ::|:|| :|||||||||||||||:||
Db   64 CPPAEAGRP----YEYYKLYLVRPEQAAACSTVLDPMVLVTCNRPEQEIRFTIKFQEFS 118

Qy   122 PNLWGHEFRSHHDYYIIATSDGTREGLESLQGGVCLLTRGMKVLLRVGQSPRGGAVPRKPV 181
                ||||| ||||:||||||||:||||:|||||||||||||:||||| ::|
Db   119 PNYMGLEFKKHHDYYITSTSNGSLEGLENREGGVCRTRTMKIIMKVGQDPNAVTPEQLTT 178

Qy   182 SEMPMPERDRGAAHSLL---PGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAGGLA--- 236
                :|::||::|| ::|||: |||||||||||||||||||||||||||||||||::
Db   179 SRPSKEADNTVKMATQAPGSRGSGLGDSDGKHETVNQEEKSGP-------GASGGSSGDPD 231

Qy   237 ---------LLLLGVAGAGGA-------------MCWRRRAKPSESRHPGPGSFGRGGSSLGL 277
                         ||||| |: |: |||||||:|:||||||||:|||||:|: ||||
Db   232 GFFNSKVALFAAVGAGCVIFLLIIFTVLLLKLPKRHRKHTQ--------RAAALSL 281

Qy   278 ----GGGGGMGPREAEPGELGIALRGGGAADPPFCPHYEKVSGDYGHPVYIVQDGPPQSP 333
                    |: ||||||:|: || |||| ||||||||||||||||||||||:|||||||| :
Db   282 STIASPRGGSGTPAGTEPSDIIIPL---FTTENNYCPHYEKVSGDYGHPVVIVQEMPPQSP 338

Qy   334 PNIYY 338
                ||||
Db   339 ANIYY 343

;

RESULT 10
US-10-138-787-13
; Sequence 13, Application US/10138787
; Patent No. US20020172984A1
; GENERAL INFORMATION:
; APPLICANT: Holland, Sacha
; APPLICANT: Mbamalu, Geraldine
; APPLICANT: Pawson, Tony
; TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED
; TITLE OF INVENTION: BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR
; TITLE OF INVENTION: TYROSINE KINASES
; FILE REFERENCE: 11757.23USWO
; CURRENT APPLICATION NUMBER: US/10/138,787
; CURRENT FILING DATE: 2002-05-03
; PRIOR APPLICATION NUMBER: US/09/214,631
; PRIOR FILING DATE: 1999-03-12
; PRIOR APPLICATION NUMBER: PCT/CA97/00473
; PRIOR FILING DATE: 1997-07-04
; PRIOR APPLICATION NUMBER: 60/021,272
; PRIOR FILING DATE: 1996-07-05
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 89
; TYPE: PRT
; ORGANISM: Homo sapiens

US-10-138-787-13

Best Local Similarity 20.0%;   Score 489; DB 9; Length 89;
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 252 RRRRAKPSESRHPGPGSPGRGGSLGLGGGGGWGPREAEPGELGIALRGGGAADPPFCPHY 311
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 RRRRAKPSESRHPGPGSPGRGGSLGLGGGGGWGPREAEPGELGIALRGGGAADPPFCPHY   60

Qy 312 EKVSGDYGHPVYIVQDGPPQSPPNIYY 338
         |||||||||||||||||||||||||||
Db   61 EKVSGDYGHPVYIVQDGPPQSPPNIYY   87

RESULT 11
US-09-862-179A-17
Sequence 17, Application US/09862179A
Patent No. US20020147306A1
GENERAL INFORMATION:
  APPLICANT: Lin, Danny
  APPLICANT: Pawson, Anthony
  TITLE OF INVENTION: PEPTIDES THAT MODULATE THE INTERACTION OF B CLASS EPHRINS
  TITLE OF INVENTION: AND PDZ DOMAINS
  FILE REFERENCE: MTSI-P01-009
  CURRENT APPLICATION NUMBER: US/09/862,179A
  CURRENT FILING DATE: 2001-05-21
  NUMBER OF SEQ ID NOS: 44
  SOFTWARE: Patentin version 3.1
  SEQ ID NO 17
  LENGTH: 89
  TYPE: PRT
  ORGANISM: Homo sapiens
US-09-862-179A-17

Query Match 100.0%;   Score 489; DB 10; Length 89;
Best Local Similarity 100.0%; Pred. No. 3.1e-29;
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 252 RRRRAKPSESRHPGPGSPGRGGSLGLGGGGGWGPREAEPGELGIALRGGGAADPPFCPHY 311
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 RRRRAKPSESRHPGPGSPGRGGSLGLGGGGGWGPREAEPGELGIALRGGGAADPPFCPHY   60

Qy 312 EKVSGDYGHPVYIVQDGPPQSPPNIYY 338
         |||||||||||||||||||||||||||
Db   61 EKVSGDYGHPVYIVQDGPPQSPPNIYY   87

RESULT 12
US-09-864-761-48262
Sequence 48262, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
  APPLICANT: Penn, Sharron G.
  APPLICANT: Rank, David R.
  APPLICANT: Hanzel, David K.
  APPLICANT: Chen, Wensheng
  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
  TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
  FILE REFERENCE: Aeomica-X-1
  CURRENT APPLICATION NUMBER: US/09/864,761
  CURRENT FILING DATE: 2001-05-23
  PRIOR APPLICATION NUMBER: US 60/180,312
  PRIOR FILING DATE: 2000-02-04
  PRIOR APPLICATION NUMBER: US 60/207,456
  PRIOR FILING DATE: 2000-05-26
  PRIOR APPLICATION NUMBER: US 09/632,366
  PRIOR FILING DATE: 2000-08-03
  PRIOR APPLICATION NUMBER: GB 24263.6
  PRIOR FILING DATE: 2000-10-04
  PRIOR APPLICATION NUMBER: US 60/236,359
  PRIOR FILING DATE: 2000-09-27
  PRIOR APPLICATION NUMBER: PCT/US01/00666

RESULT 13
US-09-864-761-48257
Sequence 48257, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
  APPLICANT: Penn, Sharron G.
  APPLICANT: Rank, David R.
  APPLICANT: Hanzel, David K.
  APPLICANT: Chen, Wensheng
  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
  TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
  FILE REFERENCE: Aeomica-X-1
  CURRENT APPLICATION NUMBER: US/09/864,761
  CURRENT FILING DATE: 2001-05-23
  PRIOR APPLICATION NUMBER: US 60/180,312
  PRIOR FILING DATE: 2000-02-04
  PRIOR APPLICATION NUMBER: US 60/207,456
  PRIOR FILING DATE: 2000-05-26
  PRIOR APPLICATION NUMBER: US 09/632,366
  PRIOR FILING DATE: 2000-08-03
  PRIOR APPLICATION NUMBER: GB 24263.6
  PRIOR FILING DATE: 2000-10-04
  PRIOR APPLICATION NUMBER: US 60/236,359
  PRIOR FILING DATE: 2000-09-27
  PRIOR APPLICATION NUMBER: US 09/774,203
  PRIOR FILING DATE: 2001-01-29
  PRIOR APPLICATION NUMBER: US 09/862,179A
  PRIOR FILING DATE: 2001-01-30
  PRIOR APPLICATION NUMBER: US 60/234,687
  PRIOR FILING DATE: 2000-09-21
  PRIOR APPLICATION NUMBER: US 09/608,408
  PRIOR FILING DATE: 2000-06-30
  PRIOR APPLICATION NUMBER: PCT/US01/00661
  PRIOR FILING DATE: 2001-01-30
  PRIOR APPLICATION NUMBER: PCT/US01/00662
  PRIOR FILING DATE: 2001-01-30
  PRIOR APPLICATION NUMBER: PCT/US01/00663
  PRIOR FILING DATE: 2001-01-30
  PRIOR APPLICATION NUMBER: PCT/US01/00664
  PRIOR FILING DATE: 2001-01-30
  PRIOR APPLICATION NUMBER: PCT/US01/00665
  PRIOR FILING DATE: 2001-01-30
  PRIOR APPLICATION NUMBER: PCT/US01/00667
  PRIOR FILING DATE: 2001-01-30
  PRIOR APPLICATION NUMBER: PCT/US01/00668
  PRIOR FILING DATE: 2001-01-30
  PRIOR APPLICATION NUMBER: PCT/US01/00669
  PRIOR FILING DATE: 2001-01-30
  PRIOR APPLICATION NUMBER: PCT/US01/00670
  PRIOR FILING DATE: 2001-01-30
  SEQ ID NO 48262
  LENGTH: 92
  TYPE: PRT
  ORGANISM: Homo sapiens
  FEATURE:
  OTHER INFORMATION: MAP TO AL136092.7
  OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.1
  OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.9
  OTHER INFORMATION: SWISSPROT HIT: P98172, EVALUE 4.00e-51
  OTHER INFORMATION: SWISSPROT HIT: P98172, EVALUE 4.00e-51
  OTHER INFORMATION: EST_HUMAN HIT: BE562822.1, EVALUE 3.00e-50
US-09-864-761-48262

Query Match 11.6%;   Score 284.5; DB 10; Length 92;
Best Local Similarity 54.2%; Pred. No. 2.6e-14;
Matches 52; Conservative 15; Mismatches 24; Indels 5; Gaps 1;

Qy   42 FQAEGGYVLYPQIGDRLDLLCPRARPPGHSSSPNYEFYKLYLYVGGAQGBRCEAPPAPNLL 101
         | :.|| :||:|||:||::|||||: ||||||||| |: |||.||: | |:||
Db   1 FLSGKGLVIYPKIGDKLDIICPRAEAGRP----YEYYKLYLYLVRPEQAAACSTVLDPNVL   55

Qy 102 LTCDRPDLDLRFTIKFQEYSPNLWGHEFRSHHDYYI 137
         :||.||: ::RFTIKFQE:SPN: G EF:.HHDYYI
Db   56 VTCNRPEQEIRFTIKFQEFSPNYMGLEFKHHDYYI   91

```
;    PRIOR FILING DATE: 2000-10-04
;    PRIOR APPLICATION NUMBER: US 60/236,359
;    PRIOR FILING DATE: 2000-09-27
;    PRIOR APPLICATION NUMBER: PCT/US01/00666
;    PRIOR FILING DATE: 2001-01-30
;    PRIOR APPLICATION NUMBER: PCT/US01/00667
;    PRIOR FILING DATE: 2001-01-30
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-297-639
;
;    PRIOR APPLICATION NUMBER: PCT/US01/00664
;    PRIOR FILING DATE: 2001-01-30
;    PRIOR APPLICATION NUMBER: PCT/US01/00669
;    PRIOR FILING DATE: 2001-01-30
;    PRIOR APPLICATION NUMBER: PCT/US01/00665
;    PRIOR FILING DATE: 2001-01-30
;    PRIOR APPLICATION NUMBER: US 60/234,687
;    PRIOR FILING DATE: 2000-09-21
;    PRIOR APPLICATION NUMBER: US 09/608,408
;    PRIOR FILING DATE: 2000-06-30
;    PRIOR APPLICATION NUMBER: US 09/774,203
;    PRIOR FILING DATE: 2001-01-29
;    PRIOR APPLICATION NUMBER: PCT/US01/00662
;    PRIOR FILING DATE: 2001-01-30
;    PRIOR APPLICATION NUMBER: PCT/US01/00663
;    PRIOR FILING DATE: 2001-01-30
;    PRIOR APPLICATION NUMBER: PCT/US01/00668
;    PRIOR FILING DATE: 2001-01-30
;    PRIOR APPLICATION NUMBER: PCT/US01/00661
;    PRIOR FILING DATE: 2001-01-30
;    PRIOR APPLICATION NUMBER: PCT/US01/00670
;    PRIOR FILING DATE: 2001-01-30
US-09-864-761-48257
; OTHER INFORMATION: SWISSPROT HIT: P52795, EVALUE 4.00e-40
; OTHER INFORMATION: EST_HUMAN HIT: AA377505.1, EVALUE 2.00e-26
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.5
; OTHER INFORMATION: MAP TO AL136092.7
; FEATURE:
; ORGANISM: Homo sapiens
; TYPE: PRT
; LENGTH: 136
SEQ ID NO 48257
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; NUMBER OF SEQ ID NOS: 49117
;    PRIOR APPLICATION NUMBER: US 09/925,297
;    PRIOR FILING DATE: 2001-01-30
```

```
Query Match              8.0%;    Score 196.5;   DB 10;   Length 136;
Best Local Similarity    36.8%;   Pred. No. 1.1e-07;
Matches    50;  Conservative   13;  Mismatches  34;  Indels  39;  Gaps   5;

Qy  230 GAAGGLA-------------LILLLGVAGAGGA-------------MCWRRRRAKPSESRHPGP 266
            |||     :                        :       ||: :
Db   11 GASGGSSGDPDGFFNSKVALFAAVGAGCVIFLLIIIIIFLIVLILKLRKRHRKHTQQ----- 65

Qy  267 GSFGRGGSLGL----GGGGGMGPREAEPGELGIALRGGGAADPPPCPHYEKVSGDYGHPV 322
            | |:         |||||||  |     |||   |||
Db   66 ----RAAALSLSTLASPKGGSGTAGTEPSDIIIPLR---TTENNYCPHYEKVSGDYGHPV 118

Qy  323 YIVQDGPPQSPPNIYY 338
            ||:| ||||| ||||
Db  119 YIVQEMPPQSPANIYY 134
```

```
RESULT 14
US-09-925-297-639
; Sequence 639, Application US/0925297
; Patent No. US20020081659A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA105
; CURRENT APPLICATION NUMBER: US/09/925,297
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05989
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
;
```

```
;    PRIOR FILING DATE: 1999-03-12
;    NUMBER OF SEQ ID NOS: 928
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 639
; LENGTH: 106
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-297-639
```

```
Query Match              7.9%;    Score 193.5;   DB 10;   Length 106;
Best Local Similarity    42.6%;   Pred. No. 1.3e-07;
Matches    46;  Conservative   11;  Mismatches  26;  Indels  25;  Gaps   4;

Qy  235 LALLLLGVAGAGGAMCWRRRRAKPSESRHPGPGSFGRGGSLGL----GGGGGMGPREAEP 290
            :   : :                                   |       |||||||
Db   18 LTVLLLKL--------------RKRHRKHTQQ-------RAAALSLSTLASPKGGSGTAGTEP 59

Qy  291 GELGIALRGGGAADPPPCPHYEKVSGDYGHPVYIVQDGPPQSPPNIYY 338
             |     |||  ||: ||||||||||||||||||: | ||||| ||||
Db   60 SDIIIPLR---TTENNYCPHYEKVSGDYGHPVYIVQEMPPQSPANIYY 104
```

```
RESULT 15
US-10-138-787-11
; Sequence 11, Application US/10138787
; Patent No. US20020129984A1
; GENERAL INFORMATION:
; APPLICANT: Holland, Sacha
; APPLICANT: Mbamalu, Geraldine
; APPLICANT: Pawson, Tony
; TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED
; TITLE OF INVENTION: BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR
; TITLE OF INVENTION: TYROSINE KINASES
; FILE REFERENCE: 11757.23USWO
; CURRENT APPLICATION NUMBER: US/10/138,787
; CURRENT FILING DATE: 2002-05-03
; PRIOR APPLICATION NUMBER: US/09/214,631
; PRIOR FILING DATE: 1999-03-12
; PRIOR APPLICATION NUMBER: PCT/CA97/00473
; PRIOR FILING DATE: 1997-07-04
; PRIOR APPLICATION NUMBER: 60/021,272
; PRIOR FILING DATE: 1996-07-05
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 82
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-138-787-11
```

```
Query Match              7.8%;    Score 192;    DB 9;    Length 82;
Best Local Similarity    46.2%;   Pred. No. 1.2e-07;
Matches    42;  Conservative    9;  Mismatches  24;  Indels  16;  Gaps   3;

Qy  252 RRRRAKPSESRHPGPGSFGRGGSLGL----GGGGGMGPREAEPGELGIALRGGGAADPPF 307
            ||                               |||||||  |     |||   ||:
Db    2 RKRHRKHTQQ-----------RAAALSLSTLASPKGGSGTAGTEPSDIIIPLR---TTENNY 49

Qy  308 CPHYEKVSGDYGHPVYIVQDGPPQSPPNIYY 338
            ||||||||||||||||||: | ||||| |||
Db   50 CPHYEKVSGDYGHPVYIVQEMPPQSPANIYY 80
```

```
RESULT 16
US-09-862-179A-15
; Sequence 15, Application US/0862179A
; Patent No. US20020147306A1
; GENERAL INFORMATION:
; APPLICANT: Lin, Danny
; APPLICANT: Pawson, Anthony
; TITLE OF INVENTION: PEPTIDES THAT MODULATE THE INTERACTION OF B CLASS EPHRINS
; TITLE OF INVENTION: AND PDZ DOMAINS
; FILE REFERENCE: MTSI-P01-009
;
```

```
;   CURRENT APPLICATION NUMBER: US/09/862,179A
;   CURRENT FILING DATE: 2001-05-21
;   NUMBER OF SEQ ID NOS: 44
;   SOFTWARE: Patentin version 3.1
;
;   SEQ ID NO 15
;   LENGTH: 82
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-862-179A-15

Query Match             7.8%;  Score 192;  DB 10;  Length 82;
Best Local Similarity   46.2%;  Pred. No. 1.2e-07;
Matches  42;  Conservative   9;  Mismatches  24;  Indels  16;  Gaps   3;

Qy    252 RRRRAKSEBSRHFGPGSFGRGGSLGL----GGGGWGPREAEPGELGIALRGGGAADPPF 307
              :  : :          : ||:::  |||||||||
Db      2 RXRHRKHTQQ-------------RAAALSLSTIASPKGSGGSGGGTTAGTEPSDIIIPLR---TTENNY 49

Qy    308 CPHYEKVSGDYGHPVVIVQDGPPQSPPNIYY 338
              ||||||||||||||||||  ||||||||
Db     50 CPHYEKVSGDYGHPVVIVQEMPPQSPANIYY 80


RESULT 17
US-10-138-787-12
;   Sequence 12, Application US/10138787
;   Patent No. US20201729841
;   GENERAL INFORMATION:
;   APPLICANT: Holland, Sacha
;   APPLICANT: Mbamalu, Geraldine
;   APPLICANT: Pawson, Tony
;   TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED
;   TITLE OF INVENTION: BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR
;   TITLE OF INVENTION: TYROSINE KINASES
;   FILE REFERENCE: 11757.23USWO
;   CURRENT APPLICATION NUMBER: US/10/138,787
;   CURRENT FILING DATE: 2002-05-03
;   PRIOR APPLICATION NUMBER: US/09/214,631
;   PRIOR FILING DATE: 1999-03-12
;   PRIOR APPLICATION NUMBER: PCT/CA97/00473
;   PRIOR FILING DATE: 1997-07-04
;   PRIOR APPLICATION NUMBER: 60/021,272
;   PRIOR FILING DATE: 1996-07-05
;   NUMBER OF SEQ ID NOS: 13
;   SOFTWARE: PatentIn Ver. 2.0
;
;   SEQ ID NO 12
;   LENGTH: 82
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-138-787-12

Query Match             7.8%;  Score 191.5;  DB 9;  Length 82;
Best Local Similarity   70.6%;  Pred. No. 1.4e-07;
Matches  36;  Conservative   4;  Mismatches   8;  Indels   3;  Gaps   1;

Qy    288 AEPGELGIALRGGGAADPPFCPHYEKVSGDYGHPVVIVQDGPPQSPPNIYY 338
              :||: :  |    |||||||||||||||||| ||||||
Db     33 SEPSDIIIPLR---TADSVFCPHYEKVSGDYGHPVVIVQEMPPQSPANIYY 80


RESULT 18
US-09-862-179A-16
;   Sequence 16, Application US/0986217A
;   Patent No. US20020147306A1
;   GENERAL INFORMATION:
;   APPLICANT: Lin, Danny
;   APPLICANT: Pawson, Anthony
;   TITLE OF INVENTION: PEPTIDES THAT MODULATE THE INTERACTION OF B CLASS EPHRINS
;   TITLE OF INVENTION: AND PDZ DOMAINS
;   FILE REFERENCE: MTSI-P01-009
;   CURRENT APPLICATION NUMBER: US/09/862,179A
;   CURRENT FILING DATE: 2001-05-21
;   NUMBER OF SEQ ID NOS: 44
;   SOFTWARE: Patentin version 3.1
;
;   SEQ ID NO 16
;   LENGTH: 82
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-862-179A-16

Query Match             7.8%;  Score 191.5;  DB 10;  Length 82;
Best Local Similarity   70.6%;  Pred. No. 1.4e-07;
Matches  36;  Conservative   4;  Mismatches   8;  Indels   3;  Gaps   1;

Qy    288 AEPGELGIALRGGGAADPPFCPHYEKVSGDYGHPVVIVQDGPPQSPPNIYY 338
              :||: :  |    |||||||||||||||||| ||||||
Db     33 SEPSDIIIPLR---TADSVFCPHYEKVSGDYGHPVVIVQEMPPQSPANIYY 80


RESULT 19
US-09-904-954-2
;   Sequence 2, Application US/09904954
;   Patent No. US20020010325A1
;   GENERAL INFORMATION:
;   APPLICANT: BECKMANN, M. P.
;   APPLICANT: CERRETTI, DOUGLAS P.
;   TITLE OF INVENTION: CYTOKINE THAT BINDS THE CELL SURFACE
;   TITLE OF INVENTION: RECEPTOR HEK
;   NUMBER OF SEQUENCES: 4
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: IMMUNEX CORPORATION
;   STREET: 51 UNIVERSITY STREET
;   CITY: SEATTLE
;   STATE: WASHINGTON
;   COUNTRY: USA
;   ZIP: 98101
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: Apple Macintosh
;   OPERATING SYSTEM: Apple System 7.1
;   SOFTWARE: Microsoft Word for Apple, Version 5.1a
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/904,954
;   FILING DATE: 12-Jul-2001
;   CLASSIFICATION: <Unknown>
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/240,124
;   FILING DATE: <Unknown>
;   APPLICATION NUMBER: US 08/114,426
;   FILING DATE: 30-AUG-1993
;   APPLICATION NUMBER: US 08/109,745
;   FILING DATE: 20-AUG-1993
;   ATTORNEY/AGENT INFORMATION:
;   NAME: SEESE, KATHRYN A.
;   REGISTRATION NUMBER: 32,172
;   REFERENCE/DOCKET NUMBER: 2814-C
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (206) 587-0430
;   TELEFAX: (206) 233-0644
;   TELEX: 756822
;   INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 238 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
;   SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-904-954-2

Query Match             7.3%;  Score 179;  DB 10;  Length 238;
Best Local Similarity   28.4%;  Pred. No. 3.8e-06;
Matches  65;  Conservative  24;  Mismatches  80;  Indels  60;  Gaps  12;

Qy      7 GPGGVRVGALLLLGVLGLVSGLSLEPVWNSANKRFQAEGGYLYPPQIGDRLDLLCP--R 64
              |||       ||| | || |: : ||::::|:::|:| |: ::| ::| ::|:
Db     24 GPG----------GALG----NRHAVVWNSSNQHLRRE-GYTVQVNNDYLDIYCPHYN 67
```

Qy   65 ARPPGPHSSP---------- -----NYEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDL----DLRFTIKF 117
Db   68 SSGVGPGAGPGPGGGAEQYVLYMVSRNGYRTCNASQGFK-RWECNRPHAPHSPIKFSEKF 126

Qy   118 QEYSPNLWGHEFRSHHDYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVP 177
Db   127 QRYSAPSLGYEFHAGHEYYYISTPTHNLH------WKCLR--MKVFVCCASTSHSG--- 174

Qy   178 RKPVSEMP---------- -----MERDRGAAHSLE------PGKENLP 204
Db   175 EKPVPTLPQFTMGPNVKINVLEDFEGENPQVPKLEKSISGTSPKREHLP 223

RESULT 20
US-09-921-984-2
Sequence 2, Application US/0921984
Patent No. US20020156239A1
GENERAL INFORMATION:
  APPLICANT: Flanagan, John G.
            Cheng, Hwai-Jong
  TITLE OF INVENTION: EPH Receptor Ligands, and Uses Related
            Thereto
  NUMBER OF SEQUENCES: 2
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: LAHIVE & COCKFIELD
    STREET: 60 State Street
    CITY: Boston
    STATE: MA
    COUNTRY: USA
    ZIP: 02109
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: ASCII(txt)
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/09/921,984
    FILING DATE: 31-Jul-2001
    CLASSIFICATION: <Unknown>
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: 08/308,814
    FILING DATE: <Unknown>
  ATTORNEY/AGENT INFORMATION:
    NAME: Vincent, Matthew P.
    REGISTRATION NUMBER: 36,709
    REFERENCE/DOCKET NUMBER: HMI-011
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: (617) 227-7400
    TELEFAX: (617) 227-5941
  INFORMATION FOR SEQ ID NO: 2:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 209 amino acids
    TYPE: amino acid
    TOPOLOGY: linear
  MOLECULE TYPE: protein
  SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-09-921-984-2

Query Match   7.2%; Score 176; DB 9; Length 209;
Best Local Similarity 29.3%; Pred. No. 5.4e-06;
Matches 58; Conservative 19; Mismatches 69; Indels 52; Gaps 7;

Qy   33 VYWNSANKRFQAE----- -GGYVLYPQIGDRLDLLCPRAPPGPHSSSPNYEFYKLYLVGGA 87
Db   35 VYWNRSNPRFQVSAVGDGCGGYTVEVSINDYLDIYCPHYGAPLP-PAERMERYILYMVNGE 93

Qy   88 QGRRCEAPPAPNLLLTCDRPDL----DLRFTIKFQEYSPNLWGHEFRSHHDYYIIATSDGT 144
Db   94 GHASCDHRQRGFKRWECNRPAAPGGPLKFSEKFQLFTPFSLGFEFRPGHEYYISATP-- 151

Qy   145 REGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKPVSEMPERDRGAAHSLEPGKENLP 204

Db   152 ----PNLVDRPCLR--LKVYVR------ -----------PTNETLY 174
Qy   205 GDP-----TSNATSRGAEG 218
Db   175 EAPEPIFISNSSCSGLGG 192

RESULT 21
US-10-138-787-7
Sequence 7, Application US/10138787
Patent No. US20020172984A1
GENERAL INFORMATION:
  APPLICANT: Mamalu, Geraldine
  APPLICANT: Pawson, Tony
  TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED
  TITLE OF INVENTION: BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR
  TITLE OF INVENTION: TYROSINE KINASES
  FILE REFERENCE: 11757.233SWO
  CURRENT APPLICATION NUMBER: US/10/138,787
  CURRENT FILING DATE: 2002-05-03
  PRIOR APPLICATION NUMBER: PCT/CA97/00473
  PRIOR FILING DATE: 1997-07-04
  PRIOR APPLICATION NUMBER: US/09/214,631
  PRIOR FILING DATE: 1999-03-12
  PRIOR APPLICATION NUMBER: 60/021,272
  PRIOR FILING DATE: 1996-07-05
  NUMBER OF SEQ ID NOS: 13
  SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 7
  LENGTH: 233
  TYPE: PRT
  ORGANISM: Homo sapiens
US-10-138-787-7

Query Match   7.1%; Score 174.5; DB 9; Length 233;
Best Local Similarity 27.4%; Pred. No. 7.8e-06;
Matches 60; Conservative 24; Mismatches 74; Indels 61; Gaps 11;

Qy   20 GVLGLIVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDLLCPR----ARPPGPHSSPN 75
Db   27 GALG-----NRHAVYWNSSNQHLRRE-GYTVQVNANDYLDIYCPHYNSSGAGPGPGGGAE 80

Qy   76 YEFYKLYLVG-----GRAQGRR---CEAPPAPNLLLTCDRPDLDLRFTIKFQEYSPNLWGH 127
Db   81 YVLYMVSRNGYRTCNASQGFKRWECNRPHAPH------SPIKFSEKFQRYSAFSLGY 131

Qy   128 EFRSHHDYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKPVSEMP-- 185
Db   132 EFHAGHEYYISTPTHNLH------WKCLR--MKVFVCCASTSHSG---EKPVPTLPQF 179

Qy   186 ------ -----MERDRGAAHSLE------PGKENLP 204
Db   180 TMGPNVKINVLEDFEGENPQVPKLEKSISGTSPKREHLP 218

RESULT 22
US-09-925-297-510
Sequence 510, Application US/09925297
Patent No. US20020016594A1
GENERAL INFORMATION:
  APPLICANT: Rosen et al.
  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
  FILE REFERENCE: PA105
  CURRENT APPLICATION NUMBER: US/09/925,297
  CURRENT FILING DATE: 2001-08-10
  PRIOR APPLICATION NUMBER: PCT/US00/05989
  PRIOR FILING DATE: 2000-03-08
  PRIOR APPLICATION NUMBER: 60/124,270
  PRIOR FILING DATE: 1999-03-12
  NUMBER OF SEQ ID NOS: 928
  SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 510

; LENGTH: 218
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-297-510

Query Match
; Best Local Similarity   28.3%; Score 172.5; DB 10; Length 218;
; Matches 51; Conservative 30; Mismatches 80; Indels 19; Gaps 6;

Qy   8 PGGVRVGALLLGVILGVLGLVSGLLSL---EPVYWNSANKRFQAEGGYVLYPQIGDRLDLLCPR 64
                                      :::::::::::::::::::::::::::
Db   8 PGQARAMEFLWAPLLGLCCSLAAADRHTVFWNSSNPKFRNE-DYIHVQLNDYVDIICPH 66

Qy   65 ARPPGFHSSPN--YEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDL---DLRFIKFQE 119
                                                :::
Db   67 YE---DHSVADAAMEQYILVLVEHEEYQLCQPQSKDQVRWQCNRPSAKHGPEKLSEKFQR 123

Qy 120 YSPNLWGHEFRSHHDYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRK 179
               ::::   :::
Db 124 FTPFTLGKEFKEGHSYYYISKPIHQHEDR------CLRLKVTVSGKITHSPQAHDNPQE 176

RESULT 23
US-09-904-954-4
; Sequence 4, Application US/0990454
; Patent No. US20020103252A1
; GENERAL INFORMATION:
; APPLICANT: BECKMANN, M. P.
; APPLICANT: CERRETTI, DOUGLAS P.
; TITLE OF INVENTION: CYTOKINE THAT BINDS THE CELL SURFACE
;   RECEPTOR HEK
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: IMMUNEX CORPORATION
; STREET: 51 UNIVERSITY STREET
; CITY: SEATTLE
; STATE: WASHINGTON
; COUNTRY: USA
; ZIP: 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Apple System 7.1
; SOFTWARE: Microsoft Word for Apple, Version 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/904,954
; FILING DATE: 12-Jul-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/240,124
; FILING DATE: <Unknown>
; APPLICATION NUMBER: US 08/114,426
; FILING DATE: 30-Aug-1993
; APPLICATION NUMBER: US 08/109,745
; FILING DATE: 20-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: SEESE, KATHRYN A.
; REGISTRATION NUMBER: 32,172
; REFERENCE/DOCKET NUMBER: 2814-C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 587-0430
; TELEFAX: (206) 233-0644
; TELEX: 756822
; TELEFAX: 756822
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 201 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-904-954-4

Query Match   6.9%; Score 169.5; DB 10; Length 201;

; Best Local Similarity   29.9%; Pred. No. 1.5e-05;
; Matches 66; Conservative 18; Mismatches 82; Indels 55; Gaps 10;

Qy   24 LVSGLLSL-EPVYWNSANKRFQAEGGYVLYPQIGDRLDLLCPRARPPPGPHSSPNYEFYKLY 82
                  ::::::::::::::::::::::::::: : ::::
Db   20 LRGGSSLRHVVYWNSSNPRL-LRGDAAVELGLNDYLDIVCPHYEGPGPPEGP--ETFALY 76

Qy   83 LVGGAQGRRCEAP-PAPNLLLTCDRPDLDLRFTIKFQEYSPNLWGHEFRSHHDYYIIATS 141
         :::   ::   ::::
Db   77 MVDWPGYESCQAEGPRAYKRWVCSLPFGHVQFSEKIQRTPFSLGFEFLPGETYYISVP 136

Qy 142 DGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKPVSEMPMERDRGAAHSLEPGKE 201
                                                        : :
Db 137 -TPB------SSGQCL------------------RLQVSVCCKERKSESAHPV----- 164

Qy 202 NLPGDPTSNATS--RGAEGPLPPPSMPAVAGAAGGLALLLL 240
                      :: :::   ::::
Db 165 ---GSPGESGTSGWRGGDTPSP---------------LCLLLL 189

RESULT 24
US-10-138-787-8
; Sequence 8, Application US/10138787
; Patent No. US20020172984A1
; GENERAL INFORMATION:
; APPLICANT: Holland, Sacha
; APPLICANT: Mbamalu, Geraldine
; APPLICANT: Pawson, Tony
; TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED
; TITLE OF INVENTION: BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR
; TITLE OF INVENTION: TYROSINE KINASES
; FILE REFERENCE: 11757.23USWO
; CURRENT APPLICATION NUMBER: US/10/138,787
; CURRENT FILING DATE: 2002-05-03
; PRIOR APPLICATION NUMBER: US/09/214,631
; PRIOR FILING DATE: 1999-03-12
; PRIOR APPLICATION NUMBER: PCT/CA97/00473
; PRIOR FILING DATE: 1997-07-04
; PRIOR APPLICATION NUMBER: 60/021,272
; PRIOR FILING DATE: 1996-07-05
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 201
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-138-787-8

Query Match
; Best Local Similarity   29.8%; Score 169; DB 9; Length 201;
; Matches 70; Conservative 19; Mismatches 84; Indels 62; Gaps 11;

Qy   17 LILGVL-----GLVSGLSL-EPVYWNSANKRFQAEGGYVLYPQIGDRLDLLCPRARPP 68
                          ::::::::::::::::::::::::::: : ::::
Db   6 LLITVLWAAFLGSPLRGGSSLRHVVYWNSSNPR-SLRGDAAVELGLNDYLDIVCPHYEGP 64

Qy   69 GPHSSPNYEFYKLYLVGGAQGRRCEAP-PAPNLLLTCDRPDLDLRFTIKFQEYSPNLWGH 127
                        :::   ::
Db   65 GPPEGP--ETFALYMVDWPGYESCQAEGPRGYXRWVCSLPFGHVQFSEKIQRFTPFSLGF 122

Qy 128 EFRSHHDYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKPVSEMPME 187
               ::::
Db 123 EFLPGETYYYISVP--TPB-----SSGQCL--------------------RLQVSVCCKE 155

Qy 188 RDRGAAHSLEPGKENLPGDPTSNATS--RGAEGPLPPPSMPAVAGAAGGLALLLL 240
                            :: :::   ::::
Db 156 RKSESAHPV------GSPGESGTSGWRGGDTPSP---------------LCLLLL 189

RESULT 25
US-10-138-787-6
; Sequence 6, Application US/10138787
; Patent No. US20020172984A1

; GENERAL INFORMATION:
; APPLICANT: Holland, Sacha
; APPLICANT: Mbamalu, Geraldine
; APPLICANT: Pawson, Tony
; TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED
; TITLE OF INVENTION: BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR
; TITLE OF INVENTION: TYROSINE KINASES
; FILE REFERENCE: 11757.23USWO
; CURRENT APPLICATION NUMBER: US/10/138,787
; CURRENT FILING DATE: 2002-05-03
; PRIOR APPLICATION NUMBER: US/09/214,631
; PRIOR FILING DATE: 1999-03-12
; PRIOR APPLICATION NUMBER: PCT/CA97/00473
; PRIOR FILING DATE: 1997-07-04
; PRIOR APPLICATION NUMBER: 60/021,272
; PRIOR FILING DATE: 1996-07-05
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patentin Ver. 2.0
;
; SEQ ID NO 6
; LENGTH: 209
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-138-787-6

Query Match   6.9%; Score 168.5; DB 9; Length 209;
Best Local Similarity   35.9%; Pred. No. 1.9e-05;
Matches   42; Conservative   14; Mismatches   52; Indels   9; Gaps   3;

Qy   33 VYWNSANKRFQAE------GRYVLYPQIGDRLDLLCPRARPPGPHSSPNYEFYKLYLVGGA 87
          |||||:| ||| :||| |:|| |||:||||||||| ::||:||:||||: ||
Db   35 VYWNRSNPRFQVSAVGDGGGVTVEVSINDYLDIYCPHYGAPLP-PAERMERYILYMVNGE 93

Qy   88 QGRRCEAPPAPNLLITCDRPDL---DLRFTIKFQEYSPNLWGHEFRSHHDYYIIATS 141
          | || |||| ||||||||||| ||||||||||||||||||:|| :||:|||||
Db   94 GHASCDHRQGFKRWECNRPAAPGGPLKFSEKFQLFTPFSLGFEFPPGHEYYYISAT 150

;
; RESULT 26
US-08-578-684-4
; Sequence 4, Application US/08578684
; Patent No. US20020137126A1
; GENERAL INFORMATION:
; APPLICANT: Caras, Ingrid W.
; APPLICANT: Winslow, John W.
; TITLE OF INVENTION: AL-1 Neurotrophic Factor
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/578,684
; FILING DATE: 02-Jan-1996
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/330128
; FILING DATE: 27-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/486449
; FILING DATE: 06-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Torchia, PhD., Timothy E.
; REGISTRATION NUMBER: 36,700
; REFERENCE/DOCKET NUMBER: P0520P2
; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 650/225-8674
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 228 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
US-08-578-684-4

Query Match   6.8%; Score 167.5; DB 8; Length 228;
Best Local Similarity   28.8%; Pred. No. 2.5e-05;
Matches   65; Conservative   29; Mismatches   81; Indels   51; Gaps   11;

Qy   33 VYWNSANKRFQAEGGYVLYPQIGDRLDLLCPRARPPGPHSSPNYEFYKLYLVG------ 85
          ||||:|| ||::|| ||::|:||:||||||||||||| ::|| :||:||||
Db   34 VYWNSSNPRFQ-RGDYHIDVCINDYLDVFCPHYEDSVPEDKI--ERYVLYMVNFDGYSAC 90

Qy   86 -----GAQGRRCEAPPAPNLLITCDRPDLDLRFTIKFQEYSPNLWGHEFRSHHDYYIIAT 140
               | | || |||| ||||||||||:|||||||||||||||||:|| :||:|||||
Db   91 DHTSKGFKRWECNRPHSEN-------GPLKFSEKFQLFTPFSLGFEFPPGREYYYISS 141

Qy   141 S---DGTREGLESLQGGVCLTRGMKVLLRVGQSPRGAAVPRKPVSEMPERDRGAAHSLE 197
              :|||| |||||||| || :|:|||||||||| ||| :| |:: || |: |||
Db   142 AIPDNGRRS--------CLK--LKVFVR-----PTNSCMKTIGVHDRVPDVNDKVENSLE 186

Qy   198 PGKBNLPGDPTSNATSRGAEGPLPPPSMPAVAGAAGGLALLLLGVA 243
          | :|| |::| :| | :||| :|:|| | | | | :|||:|||
Db   187 PADDTV--HESAEPSRG-ENAAQTPRIPSRL----LAILLFLLA 223

;
; RESULT 27
US-10-138-787-10
; Sequence 10, Application US/10138787
; Patent No. US20020172984A1
; GENERAL INFORMATION:
; APPLICANT: Holland, Sacha
; APPLICANT: Mbamalu, Geraldine
; APPLICANT: Pawson, Tony
; TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED
; TITLE OF INVENTION: BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR
; TITLE OF INVENTION: TYROSINE KINASES
; FILE REFERENCE: 11757.23USWO
; CURRENT APPLICATION NUMBER: US/10/138,787
; CURRENT FILING DATE: 2002-05-03
; PRIOR APPLICATION NUMBER: US/09/214,631
; PRIOR FILING DATE: 1999-03-12
; PRIOR APPLICATION NUMBER: PCT/CA97/00473
; PRIOR FILING DATE: 1997-07-04
; PRIOR APPLICATION NUMBER: 60/021,272
; PRIOR FILING DATE: 1996-07-05
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patentin Ver. 2.0
;
; SEQ ID NO 10
; LENGTH: 205
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-138-787-10

Query Match   6.8%; Score 166; DB 9; Length 205;
Best Local Similarity   27.5%; Pred. No. 2.8e-05;
Matches   46; Conservative   31; Mismatches   74; Indels   16; Gaps   5;

Qy   18 LLGVLGLVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDLLCPRARPPGPHSSPN-- 75
          |||: ::: :||::|||||||||||||||||||||||||||||||||||||||||||
Db   8 LIGLCCSLAAADRHTVFNNSSNFKFRNE-DYTIHVQLNDYDIICPHYE---DHSVADAA 63

Qy   76 YEFYKLYLVGGAGGRRCEAPPAPNLLITCDRPDL---DLRFTIKFQEYSPNLWGHEFRSH 132
          |:||:||||| |:||||||||||||||||||| ||||||||||||||||||||||
Db   64 MEQYILYLVEHEEYQLCQPQSKDQVRWQCNRPSAKHGPEKLSEKFQRTPFTFLGKEFKEG 123

Qy   133 HDYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRK 179
          :|||||| :||| |||||| || || :|:| |||||||| |||:|
Db   124 HSYYISKPIHQHEDR------CLRLKVTVSGKITHSPQAHVNPQE 163

RESULT 28
US-10-001-054-48
; Sequence 48, Application US/10001054
; Publication No. US20020192209A1
; GENERAL INFORMATION:
APPLICANT: Genentech, Inc.
APPLICANT: Baker, Kevin
APPLICANT: Goddard, Audrey
APPLICANT: Gurney, Austin
APPLICANT: Hebert, Carolyn
APPLICANT: Henzel, William
APPLICANT: Kabakoff, Rhona
APPLICANT: Shelton, David
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin
APPLICANT: Wood, William
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITING NEOPLASTIC
TITLE OF INVENTION: CELL GROWTH
FILE REFERENCE: P3034R1PCT
CURRENT APPLICATION NUMBER: US/10/001,054
CURRENT FILING DATE: 2001-11-30
PRIOR APPLICATION NUMBER: 60/059114
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/079689
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079920
PRIOR FILING DATE: 1998-03-30
PRIOR APPLICATION NUMBER: 60/082999
PRIOR FILING DATE: 1998-04-24
PRIOR APPLICATION NUMBER: 60/083545
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/085149
PRIOR FILING DATE: 1998-05-12
PRIOR APPLICATION NUMBER: 60/087607
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/088858
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/090691
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/100263
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/100390
PRIOR FILING DATE: 1998-09-14
PRIOR APPLICATION NUMBER: 60/096891
PRIOR FILING DATE: 1998-08-17
PRIOR APPLICATION NUMBER: 60/096894
PRIOR FILING DATE: 1998-08-17
PRIOR APPLICATION NUMBER: 60/099803
PRIOR FILING DATE: 1998-09-15
PRIOR APPLICATION NUMBER: 60/101476
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/107783
PRIOR FILING DATE: 1998-11-10
PRIOR APPLICATION NUMBER: 60/108849
PRIOR FILING DATE: 1998-11-18
PRIOR APPLICATION NUMBER: 60/112420
PRIOR FILING DATE: 1998-12-15
PRIOR APPLICATION NUMBER: 60/113296
PRIOR FILING DATE: 1998-12-22
PRIOR APPLICATION NUMBER: 60/115554
PRIOR FILING DATE: 1999-01-12
PRIOR APPLICATION NUMBER: 60/115558
PRIOR FILING DATE: 1999-01-12
PRIOR APPLICATION NUMBER: 60/116533
PRIOR FILING DATE: 1999-01-20
PRIOR APPLICATION NUMBER: 60/123618
PRIOR FILING DATE: 1999-03-10
PRIOR APPLICATION NUMBER: 60/131294
PRIOR FILING DATE: 1999-04-07
PRIOR APPLICATION NUMBER: 60/140650
PRIOR FILING DATE: 1999-06-22

PRIOR APPLICATION NUMBER: 60/141037
PRIOR FILING DATE: 1999-06-23
PRIOR APPLICATION NUMBER: 60/144758
PRIOR FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: 60/162506
PRIOR FILING DATE: 1999-07-09
PRIOR APPLICATION NUMBER: 60/170262
PRIOR FILING DATE: 1999-12-09
PRIOR APPLICATION NUMBER: 60/187202
PRIOR FILING DATE: 2000-03-03
PRIOR APPLICATION NUMBER: 60/209832
PRIOR FILING DATE: 2000-06-05
PRIOR APPLICATION NUMBER: 60/232887
PRIOR FILING DATE: 2000-09-15
PRIOR APPLICATION NUMBER: 60/180997
PRIOR FILING DATE: 1998-11-19
PRIOR APPLICATION NUMBER: 60/218517
PRIOR FILING DATE: 1998-12-22
PRIOR APPLICATION NUMBER: 09/284291
PRIOR FILING DATE: 1999-04-12
PRIOR APPLICATION NUMBER: 09/380137
PRIOR FILING DATE: 1999-08-25
PRIOR APPLICATION NUMBER: 09/380138
PRIOR FILING DATE: 1999-08-25
PRIOR APPLICATION NUMBER: 09/380913
PRIOR FILING DATE: 1999-09-09
PRIOR APPLICATION NUMBER: 09/403297
PRIOR FILING DATE: 1999-10-18
PRIOR APPLICATION NUMBER: 09/423741
PRIOR FILING DATE: 1999-11-10
PRIOR APPLICATION NUMBER: 09/709238
PRIOR FILING DATE: 2000-11-08
PRIOR APPLICATION NUMBER: 09/802706
PRIOR FILING DATE: 2001-03-09
PRIOR APPLICATION NUMBER: 09/866034
PRIOR FILING DATE: 2001-05-25
PRIOR APPLICATION NUMBER: 09/872035
PRIOR FILING DATE: 2001-06-01
PRIOR APPLICATION NUMBER: 09/882636
PRIOR FILING DATE: 2001-06-14
PRIOR APPLICATION NUMBER: 09/918585
PRIOR FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: 09/924419
PRIOR FILING DATE: 2001-08-06
PRIOR APPLICATION NUMBER: 09/927796
PRIOR FILING DATE: 2001-08-06
PRIOR APPLICATION NUMBER: 09/929404
PRIOR FILING DATE: 2001-08-13
PRIOR APPLICATION NUMBER: 09/941992
PRIOR FILING DATE: 2001-08-28
PRIOR APPLICATION NUMBER: 09/946374
PRIOR FILING DATE: 2001-09-04
PRIOR APPLICATION NUMBER: PCT/US98/18824
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: PCT/US99/00106
PRIOR FILING DATE: 1999-01-05
PRIOR APPLICATION NUMBER: PCT/US99/05028
PRIOR FILING DATE: 1999-03-08
PRIOR APPLICATION NUMBER: PCT/US99/08615
PRIOR FILING DATE: 1999-04-20
PRIOR APPLICATION NUMBER: PCT/US99/12252
PRIOR FILING DATE: 1999-06-02
PRIOR APPLICATION NUMBER: PCT/US99/20111
PRIOR FILING DATE: 1999-09-01
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28551
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/28634
PRIOR FILING DATE: 1999-12-01
PRIOR APPLICATION NUMBER: PCT/US99/30095

RESULT 29
US-10-028-072-288
Sequence 288, Application US/10028072
Publication No. US20030043111A1
GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Beresini,Maureen
; APPLICANT: DeForge,Laura
; APPLICANT: Desboyers,Luc
; APPLICANT: Filvaroff,Ellen
; APPLICANT: Gao,Wei-Qiang
; APPLICANT: Gerritsen,Mary E.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Sherwood,Steven
; APPLICANT: Smith,Victoria
; APPLICANT: Stewart,Timothy A.
; APPLICANT: Tumas,Daniel
; APPLICANT: Watanabe,Colin K
; APPLICANT: Wood,William
; APPLICANT: Zhang
; TITLE OF INVENTION:
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/028,072
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 2001-12-19
; PRIOR FILING DATE: 2000-06-18
; PRIOR APPLICATION NUMBER: 60/049911
; PRIOR FILING DATE: 1997-06-18
; PRIOR APPLICATION NUMBER: 60/056974
; PRIOR FILING DATE: 1997-08-26
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059115
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059117
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059122
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059184
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059588
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059836
; PRIOR FILING DATE: 1997-09-24
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-09-24
; PRIOR APPLICATION NUMBER: 60/062285
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/062287
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/062814
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063045
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063082
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063127
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063327
; PRIOR FILING DATE: 1997-10-27
; PRIOR APPLICATION NUMBER: 60/063329
; PRIOR FILING DATE: 1997-10-27
; PRIOR APPLICATION NUMBER: 60/063550
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063561
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063704

; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00376
; PRIOR FILING DATE: 2000-01-06
; PRIOR APPLICATION NUMBER: PCT/US00/03565
; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: PCT/US00/04341
; PRIOR FILING DATE: 2000-02-18
; PRIOR APPLICATION NUMBER: PCT/US00/04342
; PRIOR FILING DATE: 2000-02-18
; PRIOR APPLICATION NUMBER: PCT/US00/05841
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: PCT/US00/06884
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: PCT/US00/08439
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: PCT/US00/13705
; PRIOR FILING DATE: 2000-05-17
; PRIOR APPLICATION NUMBER: PCT/US00/14042
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: PCT/US00/14941
; PRIOR FILING DATE: 2000-05-30
; PRIOR APPLICATION NUMBER: PCT/US00/15264
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: PCT/US00/22031
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: PCT/US00/23522
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: PCT/US00/30873
; PRIOR FILING DATE: 2000-11-10
; PRIOR APPLICATION NUMBER: PCT/US00/32678
; PRIOR FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: PCT/US01/06520
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: PCT/US01/06666
; PRIOR FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: PCT/US01/17092
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: PCT/US01/17800
; PRIOR FILING DATE: 2001-06-01
; PRIOR APPLICATION NUMBER: PCT/US01/19692
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: PCT/US01/21066
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: PCT/US01/21735
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: PCT/US01/27099
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 91
; SEQ ID NO 48
; LENGTH: 204
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-001-054-48

Query Match   6.7%;   Score 164.5;   DB 9;   Length 204;
Best Local Similarity   27.7%;   Pred. No. 3.6e-05;
Matches 46; Conservative 31; Mismatches 74; Indels 15; Gaps   5;

Qy   18 LLGVLGLVSGLSLPPVYWNSANKRFQAEGGYVLYPQIGDRLDLLCPRARPPGPHSS-PNY 76
          |:| :: :   :||||:|:| : ||:|||: |: | :|:|: | |::|
Db   8 LLGLCCSLAAADRHTVFWNSSNPKFRNE-DYTIHVQLNDYVDIICPHYE---DHSADAAM 63

Qy   77 EFYKYLVGGAQGRRCEAPAPNLLTCDRPDL---DLRFTIKFQEYSPNLWGHEFRSHH 133
          ||| :|: | | | : | | | | :| :|: : ||:| | |:|
Db   64 EQYILYLVEHEEYQLCQPQSKDQVRWQCNRPSAKHGPEKLSEKFQRFTPFTLGKEFKEGH 123

Qy 134 DYYIIATSDGTREGLESLGGGVCLTRGMKVLLRVGQSPRGGAVPRK 179
          :||: : | || | | | ||: ||:| | | |
Db 124 SYYYSKPIHQHEDR------CLRLKVIVSGKITHSPQAHDNPQE 162

PRIOR FILING DATE: 1997-10-29
PRIOR APPLICATION NUMBER: 60/063733
PRIOR FILING DATE: 1997-10-29
PRIOR APPLICATION NUMBER: 60/063735
PRIOR FILING DATE: 1997-10-29
PRIOR APPLICATION NUMBER: 60/063738
PRIOR FILING DATE: 1997-10-29
PRIOR APPLICATION NUMBER: 60/063755
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/064248
PRIOR FILING DATE: 1997-11-03
PRIOR APPLICATION NUMBER: 60/064809
PRIOR FILING DATE: 1997-11-07
PRIOR APPLICATION NUMBER: 60/065186
PRIOR FILING DATE: 1997-11-12
PRIOR APPLICATION NUMBER: 60/065846
PRIOR FILING DATE: 1997-11-17
PRIOR APPLICATION NUMBER: 60/066364
PRIOR FILING DATE: 1997-11-21
PRIOR APPLICATION NUMBER: 60/066453
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/066511
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/066770
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/069212
PRIOR FILING DATE: 1997-12-11
PRIOR APPLICATION NUMBER: 60/069278
PRIOR FILING DATE: 1997-12-11
PRIOR APPLICATION NUMBER: 60/069334
PRIOR FILING DATE: 1997-12-11
PRIOR APPLICATION NUMBER: 60/069694
PRIOR FILING DATE: 1998-01-23
PRIOR APPLICATION NUMBER: 60/072320
PRIOR FILING DATE: 1998-02-04
PRIOR APPLICATION NUMBER: 60/073612
PRIOR FILING DATE: 1998-02-09
PRIOR APPLICATION NUMBER: 60/074086
PRIOR FILING DATE: 1998-02-09
PRIOR APPLICATION NUMBER: 60/074092
PRIOR FILING DATE: 1998-02-09
PRIOR APPLICATION NUMBER: 60/077791
PRIOR FILING DATE: 1998-03-12
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/079294
PRIOR FILING DATE: 1998-03-25
PRIOR APPLICATION NUMBER: 60/079663
PRIOR FILING DATE: 1998-02-27
PRIOR APPLICATION NUMBER: 60/079728
PRIOR FILING DATE: 1998-02-27
PRIOR APPLICATION NUMBER: 60/080165
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/081203
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/081229
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081695
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081817
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081818
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/082999
PRIOR FILING DATE: 1998-04-24
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-24
PRIOR APPLICATION NUMBER: 60/083545
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/083545
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/084600
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084627
PRIOR FILING DATE: 1998-05-07

PRIOR APPLICATION NUMBER: 60/084637
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/085149
PRIOR FILING DATE: 1998-05-12
PRIOR APPLICATION NUMBER: 60/085323
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085338
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085339
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085579
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085697
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085704
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/086414
PRIOR FILING DATE: 1998-05-22
PRIOR APPLICATION NUMBER: 60/086430
PRIOR FILING DATE: 1998-05-22
PRIOR APPLICATION NUMBER: 60/087106
PRIOR FILING DATE: 1998-05-28
PRIOR APPLICATION NUMBER: 60/088026
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088730
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088741
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088810
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088858
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/089532
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089599
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089907
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089947
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/090349
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/090429
PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/090445
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090538
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090863
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/091360
PRIOR FILING DATE: 1998-06-26
PRIOR APPLICATION NUMBER: 60/091519
PRIOR FILING DATE: 1998-07-01
PRIOR APPLICATION NUMBER: 60/091982
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/091982
PRIOR FILING DATE: 1998-07-07

```
Qy   18  LLGVLGLVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDLLCPRARPPGPHSS-PNY  76
             |||:  :    :   ||:|||:|| :  :  ||: : ||:|::    |||:  ||
Db    8  LLGLCCSLAAADRHTVFWNSSNPKFRNE-DYTIHVQLNDYVDIICPHYE---DHSADAAM  63

Qy   77  EFYKLYLVGGAQGRCEAPPAPNLLLTCDRPDL---DLRFTIKFQEYSPNLWGHEFRSHH  133
             |:|  |||  |    :  |:  |  :|| |:     |:::  :|||:|   :|
Db   64  EQYILVLVEHEEYQLCQPQSKDQVRWQCNRPSAKHGPEKLSEKFQRFTFPTLGKFFKEGH 123

Qy  134  DYYIATSDGTREGLESLGGGVCLTRGMKVLLRVGQSPRGGAVPRK  179
             |:|:   |    ||:|| |  |  :    |||  : |:  : |
Db  124  SYYYISXPIHQHEDR------CLRLKVTVSGKITHSPQAHDNPQE  162
```

Query Match   6.7%; Score 164.5; DB 9; Length 204;
Best Local Similarity 27.7%; Pred. No. 3.6e-05;
Matches 46; Conservative 31; Mismatches 74; Indels 15; Gaps 5;

RESULT 30
US-10-121-049-288
Sequence 288, Application US/10121049
Publication No. US20030022239A1
GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Beresini,Maureen
; APPLICANT: DeForge,Laura
; APPLICANT: Desnoyers,Luc
; APPLICANT: Filvaroff,Ellen
; APPLICANT: Gao,Wei-Qiang
; APPLICANT: Gerritsen,Mary E.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Sherwood,Steven
; APPLICANT: Smith,Victoria
; APPLICANT: Stewart,Timothy A.
; APPLICANT: Tumas,Daniel
; APPLICANT: Watanabe,Colin K
; APPLICANT: Wood,William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C17
; CURRENT APPLICATION NUMBER: US/10/121,049
; CURRENT FILING DATE: 2002-04-12
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 288
; LENGTH: 204
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-121-049-288

Query Match   6.7%; Score 164.5; DB 9; Length 204;
Best Local Similarity 27.7%; Pred. No.3.6e-05;
Matches 46; Conservative 31; Mismatches 74; Indels 15; Gaps   5;

Qy   18 LLGVLGLVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDLLCPRARPPGPHSS-PNY 76
        ||||:| : |||||:| |:| : |||:| :|| :|
Db   8 LLGLCCSLAAADRHTVFWNSSNPKFRNE-DYTIHVQLNDYVDIICPHYE---DHSADAAM 63

Qy   77 EFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDL---DLRFTIKFQEYSPNLWGHEFRSHH 133
        || | | |:| :| |: |||: || | | : || | |
Db   64 EQYILYLVEHEEYQLCQPQSKDQVRWQCNRPSAKGPEKLSEKFQRFTPFTLGKEFKEGH 123

Qy 134 DYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRK 179
        |: || | | ||: | | : ||: :: | |
Db 124 SYYISKPIHQHEDR-------CLRLKVTVSGKITHSPQAHDNPQE 162

RESULT 31
US-10-123-904-288
Sequence 288, Application US/10123904
Publication No. US20030022328A1
GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Beresini,Maureen
; APPLICANT: DeForge,Laura
; APPLICANT: Desnoyers,Luc
; APPLICANT: Filvaroff,Ellen
; APPLICANT: Gao,Wei-Qiang
; APPLICANT: Gerritsen,Mary E.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Sherwood,Steven
; APPLICANT: Smith,Victoria
; APPLICANT: Stewart,Timothy A.
; APPLICANT: Tumas,Daniel
; APPLICANT: Watanabe,Colin K

; APPLICANT: Wood,William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C17
; CURRENT APPLICATION NUMBER: US/10/123,904
; CURRENT FILING DATE: 2002-04-16
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 288
; LENGTH: 204
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-123-904-288

Query Match   6.7%; Score 164.5; DB 9; Length 204;
Best Local Similarity 27.7%; Pred. No.3.6e-05;
Matches 46; Conservative 31; Mismatches 74; Indels 15; Gaps   5;

Qy   18 LLGVLGLVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDLLCPRARPPGPHSS-PNY 76
        ||||:| : |||||:| |:| : |||:| :|| :|
Db   8 LLGLCCSLAAADRHTVFWNSSNPKFRNE-DYTIHVQLNDYVDIICPHYE---DHSADAAM 63

Qy   77 EFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDL---DLRFTIKFQEYSPNLWGHEFRSHH 133
        || | | |:| :| |: |||: || | | : || | |
Db   64 EQYILYLVEHEEYQLCQPQSKDQVRWQCNRPSAKGPEKLSEKFQRFTPFTLGKEFKEGH 123

Qy 134 DYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRK 179
        |: || | | ||: | | : ||: :: | |
Db 124 SYYISKPIHQHEDR-------CLRLKVTVSGKITHSPQAHDNPQE 162

RESULT 32
US-10-140-470-288
Sequence 288, Application US/10140470
Publication No. US20030022331A1
GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Beresini,Maureen
; APPLICANT: DeForge,Laura
; APPLICANT: Desnoyers,Luc
; APPLICANT: Filvaroff,Ellen
; APPLICANT: Gao,Wei-Qiang
; APPLICANT: Gerritsen,Mary E.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Sherwood,Steven
; APPLICANT: Smith,Victoria
; APPLICANT: Stewart,Timothy A.
; APPLICANT: Tumas,Daniel
; APPLICANT: Watanabe,Colin K
; APPLICANT: Wood,William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C160
; CURRENT APPLICATION NUMBER: US/10/140,470
; CURRENT FILING DATE: 2002-05-06
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 288
; LENGTH: 204
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-470-288

Query Match   6.7%; Score 164.5; DB 9; Length 204;
Best Local Similarity 27.7%; Pred. No.3.6e-05;
Matches 46; Conservative 31; Mismatches 74; Indels 15; Gaps   5;

Qy   18 LLGVLGLVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDLLCPRARPPGPHSS-PNY 76
        ||||:| : |||||:| |:| : |||:| :|| :|

Db   8 LLGLCCSLAAADRHTVFWNSSNPKFRNE-DYTIHVQLNDYVDIICPHYE---DHSADAAM 63
             |||:| :||:| ||
Qy   77 EFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDL---DLRFTIKFQEYSPNLWGHEFRSHH 133
              ||||| :||:|
Db   64 EQYILYLVEHEEYQLCQPQSKDQVRWQCNRPSAXHGPEKLSEKFQRFTPFTLGKEFKEGH 123

Qy 134 DYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRK 179
               ||:|
Db 124 SYYISKPIHQHEDR------CLRLKVTVSGKITHSPQAHDNPQE 162

RESULT 33
US-10-175-746-288
; Sequence 288, Application US/10175746
; Publication No. US20030272270A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Beresini,Maureen
; APPLICANT: DeForge,Laura
; APPLICANT: Desnoyers,Luc
; APPLICANT: Filvaroff,Ellen
; APPLICANT: Gao,Wei-Qiang
; APPLICANT: Gerritsen,Mary E.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Sherwood,Steven
; APPLICANT: Smith,Victoria
; APPLICANT: Stewart,Timothy A.
; APPLICANT: Tumas,Daniel
; APPLICANT: Watanabe,Colin K
; APPLICANT: Wood,William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P2330R1C353
; CURRENT APPLICATION NUMBER: US/10/175,746
; CURRENT FILING DATE: 2002-06-19
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 288
; LENGTH: 204
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-175-746-288

Query Match   6.7%; Score 164.5; DB 9; Length 204;
Best Local Similarity   27.7%; Pred. No.3.6e-05;
Matches 46; Conservative 31; Mismatches 74; Indels 15; Gaps 5;

RESULT 34
US-10-176-918-288
; Sequence 288, Application US/10176918
; Publication No. US20030272275A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Beresini,Maureen
; APPLICANT: DeForge,Laura
; APPLICANT: Desnoyers,Luc
; APPLICANT: Filvaroff,Ellen
; APPLICANT: Gao,Wei-Qiang
; APPLICANT: Gerritsen,Mary E.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Sherwood,Steven
; APPLICANT: Smith,Victoria
; APPLICANT: Stewart,Timothy A.
; APPLICANT: Tumas,Daniel
; APPLICANT: Watanabe,Colin K
; APPLICANT: Wood,William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P2330R1C382
; CURRENT APPLICATION NUMBER: US/10/176,918
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 288
; LENGTH: 204
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-918-288

Query Match   6.7%; Score 164.5; DB 9; Length 204;
Best Local Similarity   27.7%; Pred. No. 3.6e-05;
Matches 46; Conservative 31; Mismatches 74; Indels 15; Gaps 5;

Qy   18 LLGVLGLVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDLLCPRARPPGPHSS-PNY 76
             |:| ::   :|: :||| || ::
Db   8 LLGLCCSLAAADRHTVFWNSSNPKFRNE-DYTIHVQLNDYVDIICPHYE---DHSADAAM 63

Qy   77 EFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDL---DLRFTIKFQEYSPNLWGHEFRSHH 133
              |||:| :||:| ||
Db   64 EQYILYLVEHEEYQLCQPQSKDQVRWQCNRPSAXHGPEKLSEKFQRFTPFTLGKEFKEGH 123

Qy 134 DYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRK 179
               ||:|
Db 124 SYYISKPIHQHEDR------CLRLKVTVSGKITHSPQAHDNPQE 162

RESULT 35
US-10-176-921-288
; Sequence 288, Application US/10176921
; Publication No. US20030272276A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Beresini,Maureen
; APPLICANT: DeForge,Laura
; APPLICANT: Desnoyers,Luc
; APPLICANT: Filvaroff,Ellen
; APPLICANT: Gao,Wei-Qiang
; APPLICANT: Goddard,Audrey
; APPLICANT: Gerritsen,Mary E.
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Sherwood,Steven
; APPLICANT: Smith,Victoria
; APPLICANT: Stewart,Timothy A.
; APPLICANT: Tumas,Daniel
; APPLICANT: Watanabe,Colin K
; APPLICANT: Wood,William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P2330R1C288
; CURRENT APPLICATION NUMBER: US/10/176,921
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 288
; LENGTH: 204

```
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-921-288

Query Match      6.7%;    Score 164.5;  DB 9;  Length 204;
Best Local Similarity  27.7%;  Pred. No. 3.6e-05;
Matches  46;  Conservative  31;  Mismatches  74;  Indels  15;  Gaps  5;

Qy   18  LLGVGLVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDLLCPRARPPGHSS-PNY   76
             |:|||||  |:  |:  |  |::   |  |:|  |  ||  |  |||:
Db    8  LLGLCCSLAAADRHTVFWNSSNPKFRNE-DYTIHVQLNDVDIICPHYE---DHSADAM   63

Qy   77  EFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDL--DLRFTIKFQEYSPNLWGHEFRSHH  133
             | ||  |||  |      ||    |  |:|     |: |:  |   | || || :
Db   64  EQYILYLVEHEEYQLCQPQSKDQVRWQCNRPSAKGPEKLSEKFQRFTFTLGKEFKGH   123

Qy  134  DYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRK  179
             |:|                            ||:    ::
Db  124  SYYYISKPIHQHEDR-------CLRLKVYIVSGKITHSPQAHDNPQE  162
```

```
RESULT 36
US-10-138-787-9
; Sequence 9, Application US/10138787
; Patent No. US2002017298A1
; GENERAL INFORMATION:
; APPLICANT: Holland, Sacha
; APPLICANT: Mbamalu, Geraldine
; APPLICANT: Pawson, Tony
; TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED
; TITLE OF INVENTION: BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR
; TITLE OF INVENTION: TYROSINE KINASES
; FILE REFERENCE: 11757.23USNO
; CURRENT APPLICATION NUMBER: US/10/138,787
; CURRENT FILING DATE: 2002-05-03
; PRIOR APPLICATION NUMBER: US/09/214,631
; PRIOR FILING DATE: 1999-03-12
; PRIOR APPLICATION NUMBER: PCT/CA97/00473
; PRIOR FILING DATE: 1997-07-04
; PRIOR APPLICATION NUMBER: 60/021,272
; PRIOR FILING DATE: 1996-07-05
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 9
; LENGTH: 228
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-138-787-9
```

```
Query Match      6.6%;    Score 160.5;  DB 9;  Length 228;
Best Local Similarity  28.3%;  Pred. No. 8e-05;
Matches  64;  Conservative  29;  Mismatches  82;  Indels  51;  Gaps  11;

Qy   33  VYWNSANKRFQAEGGYVLYPQIGDRLDLLCPRARPPGPHSSSPNYEFYKLYLYG------   85
             ||||  |   |:|
Db   34  VYWNSSNPKFQ-RDYHIDVCINDYLDVFCPHYEDSVPEDKT--ERYLYMVNPDGYSAC    90

Qy   86  -----GAQGRRCEAPPAPNLLLTCDRPDLDLRFTIKFQEYSPNLWGHEFRSHHDYYIIAT  140
Db   91  DHTSKGFKRWECNRPHSPN-----------GPLKFSEKFQLFTPFSLGFFPPGREYFYISS  141

Qy  141  S---DGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKPVPVSEMPMERDRGAAHSLE  197
Db  142  AIPDNGRRS-----------CLK--LKVFVR-----PTNSCMKTIGVHDRVFDVANDKVENSLE  186

Qy  198  PGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAGGLALLLLGVA  243
Db  187  PADDDV--HESAEPSRG-ENAAQTPRIPSRL-----LAILLFLLA  223
```

```
RESULT 37
US-09-925-301-1225
; Sequence 1225, Application US/09925301
```

```
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1225
; LENGTH: 27
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-301-1225

Query Match      5.9%;    Score 145;  DB 10;  Length 27;
Best Local Similarity  100.0%;  Pred. No. 9.2e-05;
Matches  25;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy  314  VSGDYGHPVYIVQDGPPQSPPNIYY  338
             |||||||||||||||||||||||||
Db    1  VSGDYGHPVYIVQDGPPQSPPNIYY  25
```

```
RESULT 38
US-09-862-179A-2
; Sequence 2, Application US/09862179A
; Patent No. US20020147306A1
; GENERAL INFORMATION:
; APPLICANT: Lin, Danny
; APPLICANT: Pawson, Anthony
; TITLE OF INVENTION: PEPTIDES THAT MODULATE THE INTERACTION OF B CLASS EPHRINS
; TITLE OF INVENTION: AND PDZ DOMAINS
; FILE REFERENCE: MTSI-P01-009
; CURRENT APPLICATION NUMBER: US/09/862,179A
; CURRENT FILING DATE: 2001-05-21
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 2
; LENGTH: 28
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: PDZ inhibitory peptide
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (20)..(20)
; OTHER INFORMATION: Xaa=Glu or Asp
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (21)..(21)
; OTHER INFORMATION: Xaa=Met or Gly
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (27)..(27)
; OTHER INFORMATION: Xaa=Ala or Pro
US-09-862-179A-2

Query Match      5.8%;    Score 141;  DB 10;  Length 28;
Best Local Similarity  92.3%;  Pred. No. 0.00019;
Matches  24;  Conservative  0;  Mismatches  2;  Indels  0;  Gaps  0;

Qy  308  CPHYEKVSGDYGHPVYIVQDGPPQSP  333
             ||||||||||||||||||||||||
Db    1  CPHYEKVSGDYGHPVYIVQXXPPQSP  26
```

```
RESULT 39
US-10-001-887-108
```

```
;  Sequence 108, Application US/10001887
;  Patent No. US20020155464A1
;  GENERAL INFORMATION:
;  APPLICANT: Salceda, Susana
;  APPLICANT: Macina, Roberto
;  APPLICANT: Recipon, Herve
;  APPLICANT: Cafferkey, Robert
;  APPLICANT: Sun, Yongming
;  APPLICANT: Liu, Chenghua
;  TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and Pr
;  FILE REFERENCE: DEX-0269
;  CURRENT APPLICATION NUMBER: US/10/001,887
;  CURRENT FILING DATE: 2001-11-20
;  PRIOR APPLICATION NUMBER: 60/249,998
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: 60/252,563
;  PRIOR FILING DATE: 2000-11-22
;  NUMBER OF SEQ ID NOS: 137
;  SOFTWARE: PatentIn version 3.1
;  SEQ ID NO 108
;  LENGTH: 638
;  TYPE: PRT
;  ORGANISM: Homo sapien
US-10-001-887-108

Query Match             5.7%;
Best Local Similarity   26.4%;
Matches   97;  Conservative  20;  Mismatches 136;  Indels 115;  Gaps  18;

Qy     2 GPPHSGPGGVRV-GAILLLGVLGLVSGLSLEPVYWNSANKRPQABGGYVLYPQIGDRLDL  60
Db   208 GPP--GVDGVGVPGAAGLPGPQG-PSGAKGEP--------------------GTR---  239

Qy    61 LCPRARPPGPHSSPNYEFYKLYLVGGAQGRRCEAPPA--PNLLLTCDRPDLDLRFIKFQ 118
Db   240 ----GPPGLIGPTGYG------MPGLPGPKGDRGPAGVPGLLGDRGEPGED-------  280

Qy   119 EVSPNLWGHEFRSHHDYIIATSDGTREGLESLQGGVCLTRGMKVLLRVG-QSPRGGAVP 177
Db   281 -------------GDPGEQCPQGLGGPPGCLPGSAGLPGRRGPPGPFKGEAGP  318

Qy   178 RKPVSEMPMERDRG----AAHSLEPGKENLPGPTSNATSRGABGPLPPPSMPAVAGAAG 233
Db   319 GGPPGVGIRGDQGPSGLAGKPGVPGERGLPG---------AHGPPGTGPKGEPGFTGRPG 371

Qy   234 GLALILLGVAGAGGAMCWRRRAKPSESRHPG-PGSFGRGGSLGL-GGGGGMGPR----- 286
Db   372 G------PGVAGALG------QKGDLGLPGQPGLRGPSGIPGLQSPAGPIGPQGLPGL  417

Qy   287 EABPGELGIALRG-------GGAADPPFCPHYEKVSGDYGHPFVYIVQDGPPQSPPNIYYT 339
Db   418 KGEPGLPGPPGEGRGAEPGTAGPTGPVGPVPGSPGITGPPGPP---GPPGPPGPAPGAFDET 474

Qy   340 SISVLEWP 347
Db   475 GIAGLHLP 482
```

```
RESULT 40
US-09-823-240-10
;  Sequence 10, Application US/09823240
;  Patent No. US20020048813A1
;  GENERAL INFORMATION:
;  APPLICANT: Frank B. Gertler
;  APPLICANT: James E. Bear
;  APPLICANT: Jurgen Wehland
;  APPLICANT: Joseph Loureiro
;  TITLE OF INVENTION: Methods and Products for Regulating Cell
;  TITLE OF INVENTION: Motility
;  FILE REFERENCE: M0656/7064 (HCL)
;  CURRENT APPLICATION NUMBER: US/09/823,240
;  CURRENT FILING DATE: 2001-03-30
;  PRIOR APPLICATION NUMBER: 60/194,564
```

```
;  PRIOR FILING DATE: 2000-04-03
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: FastSEQ for Windows Version 3.0
;  SEQ ID NO 10
;  LENGTH: 380
;  TYPE: PRT
;  ORGANISM: Homo sapien
US-09-823-240-10

Query Match             5.3%;
Best Local Similarity   22.8%;
Matches   67;  Conservative  34;  Mismatches 102;  Indels  91;  Gaps  15;

Qy    35 WNSANKRPQABGGYVLYPQIGDRLDLLCPRARPPGPHSSPNYEFYKLYLVGGAQGRRCEA  94
Db    16 YDDGNKRWLPAG---TGPQAFSRVQI--------------YHNPTANSPRVV------GRKMQ-  55

Qy    95 PPAPNLLLTCDRPDLDLRFIKFQEYSPN----------IWGHEFRSHHDYIIATSDGTR 145
Db    56 -PDQQVVINC-----AIVRGVKYNQATPNFHQWRDARQVWGNLNFGGSKEDAAQPAA--GMA 107

Qy   146 EGLESLQGGVCLTRGMKVLLRVGVGQSPRGG---AVPRKPVYSEMPMERDR---GAAHSLE- 197
Db   108 SALEALEGG------------GPPPPPALPTWSVPNGPSPEEVEQQKROQPGPSEHIER 154

Qy   198 -----------------------------PGKENLPGDPTSN--ATSRGA-EGPLPPPSMPAVA 229
Db   155 RVSNAGGPPAPPAGPPGPPPGPPPGPPGGPLPPSGVPAAAHGAGGGPPPAPPLPAAQ 214

Qy   230 GAAGGLALILLGVAGAGGAMCWRRRAKPSESRHPGPGSFGRGGSLGLGLGGGGM 283
Db   215 GPGGGGA-----GAPGLAAAIAGAKLR-KVSKOEEASGGGPTAPKAESGRSGGGGL 263
```

```
RESULT 41
US-09-925-299-979
;  Sequence 979, Application US/09925299
;  Patent No. U82002005627A1
;  GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA102
;  CURRENT APPLICATION NUMBER: US/09/925,299
;  CURRENT FILING DATE: 2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05883
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 1556
;  SOFTWARE: PatentIn Ver. 2.0
;  SEQ ID NO 979
;  LENGTH: 674
;  TYPE: PRT
;  ORGANISM: Homo sapiens
US-09-925-299-979

Query Match             5.3%;
Best Local Similarity   27.4%;
Matches   72;  Conservative  20;  Mismatches 92;  Indels  79;  Gaps  15;

Qy   141 SDGTREGLESLQGGVCLTRGMKVLLRVGQSP--RG--GAVPRKPVYSEMPME--RDRGAAH 194
Db   186 SRGER-GLPGVAGAV----GEPGPLGIAGPPGARGPPGAVGSPGVNGAPGEBAGRDGNPGN 240

Qy   195 SLEPGKENLPG----------DPTSNATSRGABGPLPP---------------PSMPAVAGAA 232
Db   241 DGPPGRDGQPGHKGERGYPGNIGVGVAAGAPGEHGPVGPAGKHGNRGETGPSGP--VGPA 298

Qy   233 GGLALILLGVAGAGGAMCWRRRAKPSESRHPG-PGSFGRGGSLGL--------GGGG 281
Db   299 GA------VGPRGPSGPGGIRGDKGEPGEKGPRGLPGLLKGHNGLQGLPGIIAGHHGDGAPG 353

Qy   282 GMGP--------------------REAEPGELGIA-LRG-------GGAADPPFCPHYEKVSG 316
```

Db   354 SVGPAGPRGPRGPAGPSGPAGKDGRTGHPGTVGPAGPGPGPGPPGGPPGVSG 413
          :|| |||||:|| |||
Qy   317 ---DYGHPVIVQDGPPQSPPNI 336

Db   414 GGYDFGYGDFYRADQPRSAPSI 436

RESULT 42
US-09-854-133-187
; Sequence 187, Application US/09854133
; Publication No. US20020183499A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Henderson, Robert A.
; APPLICANT: Benson, Darin R.
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C10
; CURRENT APPLICATION NUMBER: US/09/854,133
; CURRENT FILING DATE: 2001-05-11
; NUMBER OF SEQ ID NOS: 735
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 187
; LENGTH: 595
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-854-133-187

Query Match   5.3%; Score 129; DB 9; Length 595;
Best Local Similarity   27.4%; Pred. No. 0.048;
Matches   69; Conservative 18; Mismatches 89; Indels 76; Gaps 12;

Qy   85 GGAGG-RRCEAPPAPNLLLTCDRPDLDLRFTIKFQEYSPNLWGH------------- 127
           :|:| |:| ||:|:|:|
Db   163 GGAESPRGWEAGP----------------REWGPSPSGHGDGPRRPRKRRGRKG 201

Qy   128 EFRSHHDYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKPVSE---- 183
           :|:|:|:| |||||||| |||
Db   202 RMGRQHEAAATAATAATATGGTAEEAGASAPES-----QAGGGPRGRA--RGPRQQGRRR 254

Qy   184 MPMERDRGAAHSLEPGKENLPGDPTS---NATSRGAEGPLPPPSMPAVAGAAGGLALLLL 240
           :|:|:||||||| ::||:|:|||
Db   255 HGTQRRRGPPQAREEG-----PRDATTILGLGTPSGEQRADQSQALPALAGAAAAHAHAIP 310

Qy   241 G---------VAGAGGAMCWRRRRAKPSESRHPGPGSFGRGGSLGLGGGGGMGPREAEPGEL 293
                       |||||||
Db   311 GAGPAAAPVGRGRRGGWRGGRRGGS-----AGAGGGRGGR-GRGRGGGRGG-------- 357

Qy   294 GIALRGGGAADP 305
           |||||||||
Db   358 GGAGRGGGAAGP 369

RESULT 43
US-09-738-973-187
; Sequence 187, Application US/09738973
; Patent No. US20020110563A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Fling, Steven P.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Algate, Paul A.
; APPLICANT: Secrist, Heather
; APPLICANT: Indrias, Carol Yoseph
; APPLICANT: Benson, Darin R.
; APPLICANT: Elliot, Mark
; APPLICANT: Mannion, Jane
; APPLICANT: Kalos, Michael D.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C9
; CURRENT APPLICATION NUMBER: US/09/738,973
; CURRENT FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 587
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 187
; LENGTH: 595
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-738-973-187

Query Match   5.3%; Score 129; DB 10; Length 595;
Best Local Similarity   27.4%; Pred. No. 0.048;
Matches   69; Conservative 18; Mismatches 89; Indels 76; Gaps 12;

Qy   85 GGAGG-RRCEAPPAPNLLLTCDRPDLDLRFTIKFQEYSPNLWGH------------- 127
           :|:| |:| ||:|:|:|
Db   163 GGAESPRGWEAGP----------------REWGPSPSGHGDGPRRPRKRRGRKG 201

Qy   128 EFRSHHDYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKPVSE---- 183
           :|:|:|:| |||||||| |||
Db   202 RMGRQHEAAATAATAATATGGTAEEAGASAPES-----QAGGGPRGRA--RGPRQQGRRR 254

Qy   184 MPMERDRGAAHSLEPGKENLPGDPTS---NATSRGAEGPLPPPSMPAVAGAAGGLALLLL 240
           :|:|:||||||| ::||:|:|||
Db   255 HGTQRRRGPPQAREEG-----PRDATTILGLGTPSGEQRADQSQALPALAGAAAAHAHAIP 310

Qy   241 G---------VAGAGGAMCWRRRRAKPSESRHPGPGSFGRGGSLGLGGGGGMGPREAEPGEL 293
                       |||||||
Db   311 GAGPAAAPVGRGRRGGWRGGRRGGS-----AGAGGGRGGR-GRGRGGGRGG-------- 357

Qy   294 GIALRGGGAADP 305
           |||||||||
Db   358 GGAGRGGGAAGP 369

RESULT 44
US-09-919-497-56
; Sequence 56, Application US/09919497
; Patent No. US20020106662A1
; GENERAL INFORMATION:
; APPLICANT: Mutter, George L.
; TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER
; FILE REFERENCE: B0801/7225
; CURRENT APPLICATION NUMBER: US/09/919,497
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/221,735
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 56
; LENGTH: 1806
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (758)..(758)
; OTHER INFORMATION: Xaa = any amino acid
; NAME/KEY: UNSURE
; LOCATION: (809)..(809)
; OTHER INFORMATION: Xaa = any amino acid
US-09-919-497-56

Query Match   5.3%; Score 129; DB 10; Length 1806;
Best Local Similarity   29.1%; Pred. No. 0.17;
Matches   64; Conservative 11; Mismatches 65; Indels 80; Gaps 13;

Qy   169 QSPRGGAVPRKPVSEMPMER------------DRGAAHSLEPGKP---------NLPGD-------- 206
           ||||| :||||:| |||:|||| |||
Db   561 QGPRGVGQPPGPTGK-PGKRGRPGADGGRGMPGPGAKGDRGFDGLPGLPGDKGHRGERG 619

```
Qy  207 ------PTSNATSRGAEGPLPPPSMPAVAGAAGGIALLLLGVAGAGGAMCWRRRAKPSE 260
Db  620 PQGPGPGPGPDDGWRGEDGEIGPRGLPGEAGPRG------LLGPRGTPGA------PGQ 665

Qy  261 ------SRHPG-PGSFGRGGSLGL-GGGGGMGGPREAEPGELGIALRGG 300
Db  666 PGMAGVDGPGPGPKGNMGPGGEGPGPPGQGGNPGPGLPGPQGPIGP----PGHGKPGGKPG 721

Qy  301 -----GAADPPFCPHYEKVSGDYGHPVVIVQDGP--PQSP 333
Db  722 LAGLPGADGPPGHPGKEGQSGEKG------------ALGPPGPGQGP 755
```

RESULT 45
US-09-925-302-518
; Sequence 518, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 518
; LENGTH: 430
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (11)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-302-518

Query Match   5.2%;   Score 127; DB 10; Length 430;
Best Local Similarity 29.3%;   Pred. No. 0.046;
Matches 49; Conservative 13; Mismatches 69; Indels 36; Gaps 6;

```
Qy  198 PGKENLPGD------------------PTSNATSRGAEGPLPPPSMPAVAGAAG--GIALLLLG 241
            |||||||                          |:|
Db   55 PGVFGLPGEKGPRGEQGEMGNIGPTGAVGDRGPEKGKGDPGPFPGAPGTVGAPGTGIAGTPQK 114

Qy  242 VAGAGGAMCWRRRAKPSESRHPGPGSPGRGGSLGLGGGGGMGGPREAEPGELGIALRGG 301
Db  115 IAVQPGTVGAPGGRGPG------APGEMGPQGPGEPGFRG------APGKAGPQGRGGV 163

Qy  302 AADPPFCPHYEKVSGDYGHPVVIVQDGPPQSP-----PNIYTISISV 343
Db  164 SAVPGF----RGDEGPIGHQGPIGQEGAPGRGSPGLPGMPGRSVSI 206
```

RESULT 46
US-09-823-240-9
; Sequence 9, Application US/09823240
; Patent No. US20020048813A1
; GENERAL INFORMATION:
; APPLICANT: Frank B. Gertler
; APPLICANT: James E. Bear
; APPLICANT: Jurgen Wehland
; APPLICANT: Joseph Loureio
; TITLE OF INVENTION: Methods and Products for Regulating Cell
; TITLE OF INVENTION: Motility
; FILE REFERENCE: M0656/7064 (HCL)
; CURRENT APPLICATION NUMBER: US/09/823,240
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 60/194,564
; PRIOR FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 684
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-823-240-9

Query Match   5.2%;   Score 127; DB 10; Length 684;
Best Local Similarity 21.5%;   Pred. No. 0.079;
Matches 82; Conservative 34; Mismatches 151; Indels 115; Gaps 15;

```
Qy   67 PPGPHSSPNYEFYKLYLVGGAQGRCEAPAPAPNLLLTCDRRPDLDLRFTIKPQEY------ 120
Db  234 PPAPQQQPQOQQFQOQOQOQOQYQQMVQAGYAPS--------------QQYQQPHYV 275

Qy  121 -----SPNLWGHEFRSHHDYIYIIATSDGTRSGLESLQGGVCLTRGMKVLLRVGQSPRGGAV 176
Db  276 LSNSNPNLTVHIQYPTQQ-----AQOQPPQAPQPPLONG-----GMYM---VGHSRLPSSA 322

Qy  177 PRKPVSEMPMERDRGAAHSLEPGKENLPG---------DPTSNATSRGAEGPLPPPSMPA 227
Db  323 SANSVVYASQQOMLPQAHPQAPTMPGGYGGPPVPPQQQOAENPYGQVPMPPPMNFS 382

Qy  228 VAGAGGLALLLLGVAGAGGAMCWRRRAKPSESRHPGPGSPGRGGSLGLGGGG----- 281
Db  383 QQQOPQGVPLNRMSSQGGPG--------GPPAPAPPPPPPSFG----GAAGGGPPPAP 429

Qy  282 -------GMGPREAEPGELGIALRGGGAADPPFCPHYEKVSGDYGHPVVIVQDG 328
Db  430 PQMFNGAPPPPAMGGGPPPAPPAPPAM---GGG----PPPAP-----GGPGAP----- 471

Qy  329 PPQSPENIYYTSISVLEWPILHTIQLFFMVRSKCSRVTTFLFPVQVITT----------STCR 380
Db  472 PPPPPPGLGGA------PKKEDPQADLMGSLASQLQQFKLKKRKVTTSAPENSGSSTSS 524

Qy  381 MTSFSFTTLNPSMQACRAQMGE 402
Db  525 GGSGNYGTIGRSSNGMVASMMDE 546
```

RESULT 47
US-09-925-302-689
; Sequence 689, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 689
; LENGTH: 403
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (183)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-302-689

Query Match   5.2%;   Score 126.5; DB 10; Length 403;
Best Local Similarity 30.9%;   Pred. No. 0.046;
Matches 46; Conservative 9; Mismatches 53; Indels 41; Gaps 8;

```
Qy  198 PGKENLPGDPTSNATS-----RGAEGPLPPPSMPAVAGAAGGLALLLLGVAGAGGAMCWR 252
Db  168 PGAPGPGPGPPAEKGAXGAMGRDGATGPSGPGPGQGVKGEAG-----LQGPQGAPG---- 217
```

```
Qy    253  RRRAKPSESRHPGP----GSFGRGGSLGLGGGGGMGPREAEPGELGI--------ALRGGG  301
                                                                        
Db    218  ----KQGATGTPGPGQSEKGSKGDGGLIGPXGETG---TKGEKGDLGLPGSSKGDRGMKGDA  270


Qy    302  AADPPFCPHYEKVSGDYGHPYVIVQDGPP  330
                                                 
Db    271  GVMGP--PGAQGSKGDFGRP------GPP  291


RESULT 48
US-09-978-295A-614
Sequence 614, Application US/09978295A
Patent No. US20020156006A1
GENERAL INFORMATION:
    APPLICANT: Ashkenazi, Avi
    APPLICANT: Baker Kevin P.
    APPLICANT: Botstein, David
    APPLICANT: Desnoyers, Luc
    APPLICANT: Baton, Dan
    APPLICANT: Ferrara, Napoleon
    APPLICANT: Filvaroff, Ellen
    APPLICANT: Fong, Sherman
    APPLICANT: Gao, Wei-Qiang
    APPLICANT: Gerber, Hanspeter
    APPLICANT: Gerritsen, Mary E.
    APPLICANT: Goddard, Audrey
    APPLICANT: Godowski, Paul J.
    APPLICANT: Grimaldi, J. Christopher
    APPLICANT: Gurney, Austin L.
    APPLICANT: Hilian, Kenneth J
    APPLICANT: Kljavin, Ivar J.
    APPLICANT: Kuo, Sophia S.
    APPLICANT: Napier, Mary A.
    APPLICANT: Pan, James;
    APPLICANT: Paoni, Nicholas F.
    APPLICANT: Roy, Margaret Ann
    APPLICANT: Shelton, David L.
    APPLICANT: Stewart, Timothy A.
    APPLICANT: Tumas, Daniel
    APPLICANT: Williams, P. Mickey
    APPLICANT: Wood, William I.
    TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
    TITLE OF INVENTION: Acids Encoding the Same
    FILE REFERENCE: P2630P1C11
    CURRENT APPLICATION NUMBER: US/09/978,295A
    CURRENT FILING DATE: 2001-10-15
    PRIOR APPLICATION NUMBER: 09/918585
    PRIOR FILING DATE: 2001-07-30
    PRIOR APPLICATION NUMBER: 60/062250
    PRIOR FILING DATE: 1997-10-17
    PRIOR APPLICATION NUMBER: 60/064249
    PRIOR FILING DATE: 1997-11-03
    PRIOR APPLICATION NUMBER: 60/065311
    PRIOR FILING DATE: 1997-11-13
    PRIOR APPLICATION NUMBER: 60/066364
    PRIOR FILING DATE: 1997-11-21
    PRIOR APPLICATION NUMBER: 60/077450
    PRIOR FILING DATE: 1998-03-10
    PRIOR APPLICATION NUMBER: 60/077632
    PRIOR FILING DATE: 1998-03-11
    PRIOR APPLICATION NUMBER: 60/077641
    PRIOR FILING DATE: 1998-03-11
    PRIOR APPLICATION NUMBER: 60/077649
    PRIOR FILING DATE: 1998-03-11
    PRIOR APPLICATION NUMBER: 60/077791
    PRIOR FILING DATE: 1998-03-12
    PRIOR APPLICATION NUMBER: 60/078004
    PRIOR FILING DATE: 1998-03-13
    PRIOR APPLICATION NUMBER: 60/078886
    PRIOR FILING DATE: 1998-03-20
    PRIOR APPLICATION NUMBER: 60/078936
    PRIOR FILING DATE: 1998-03-20
```

```
    PRIOR APPLICATION NUMBER: 60/078910
    PRIOR FILING DATE: 1998-03-20
    PRIOR APPLICATION NUMBER: 60/078939
    PRIOR FILING DATE: 1998-03-20
    PRIOR APPLICATION NUMBER: 60/079663
    PRIOR FILING DATE: 1998-03-27
    PRIOR APPLICATION NUMBER: 60/079689
    PRIOR FILING DATE: 1998-03-27
    PRIOR APPLICATION NUMBER: 60/079728
    PRIOR FILING DATE: 1998-03-27
    PRIOR APPLICATION NUMBER: 60/079786
    PRIOR FILING DATE: 1998-03-25
    PRIOR APPLICATION NUMBER: 60/079920
    PRIOR FILING DATE: 1998-03-30
    PRIOR APPLICATION NUMBER: 60/079923
    PRIOR FILING DATE: 1998-03-30
    PRIOR APPLICATION NUMBER: 60/080105
    PRIOR FILING DATE: 1998-03-31
    PRIOR APPLICATION NUMBER: 60/080107
    PRIOR FILING DATE: 1998-03-31
    PRIOR APPLICATION NUMBER: 60/080165
    PRIOR FILING DATE: 1998-03-31
    PRIOR APPLICATION NUMBER: 60/080194
    PRIOR FILING DATE: 1998-03-31
    PRIOR APPLICATION NUMBER: 60/080327
    PRIOR FILING DATE: 1998-04-01
    PRIOR APPLICATION NUMBER: 60/080334
    PRIOR FILING DATE: 1998-04-01
    PRIOR APPLICATION NUMBER: 60/080070
    PRIOR FILING DATE: 1998-04-01
    PRIOR APPLICATION NUMBER: 60/080328
    PRIOR FILING DATE: 1998-04-01
    PRIOR APPLICATION NUMBER: 60/080333
    PRIOR FILING DATE: 1998-04-01
    PRIOR APPLICATION NUMBER: 60/081049
    PRIOR FILING DATE: 1998-04-08
    PRIOR APPLICATION NUMBER: 60/081071
    PRIOR FILING DATE: 1998-04-08
    PRIOR APPLICATION NUMBER: 60/081195
    PRIOR FILING DATE: 1998-04-08
    PRIOR APPLICATION NUMBER: 60/081203
    PRIOR FILING DATE: 1998-04-09
    PRIOR APPLICATION NUMBER: 60/081229
    PRIOR FILING DATE: 1998-04-09
    PRIOR APPLICATION NUMBER: 60/081955
    PRIOR FILING DATE: 1998-04-15
    PRIOR APPLICATION NUMBER: 60/081817
    PRIOR FILING DATE: 1998-04-15
    PRIOR APPLICATION NUMBER: 60/081819
    PRIOR FILING DATE: 1998-04-15
    PRIOR APPLICATION NUMBER: 60/081952
    PRIOR FILING DATE: 1998-04-15
    PRIOR APPLICATION NUMBER: 60/081838
    PRIOR FILING DATE: 1998-04-15
    PRIOR APPLICATION NUMBER: 60/082569
    PRIOR FILING DATE: 1998-04-21
    PRIOR APPLICATION NUMBER: 60/082568
    PRIOR FILING DATE: 1998-04-21
    PRIOR APPLICATION NUMBER: 60/082704
    PRIOR FILING DATE: 1998-04-22
    PRIOR APPLICATION NUMBER: 60/082804
    PRIOR FILING DATE: 1998-04-22
    PRIOR APPLICATION NUMBER: 60/082700
    PRIOR FILING DATE: 1998-04-22
    PRIOR APPLICATION NUMBER: 60/082797
    PRIOR FILING DATE: 1998-04-22
    PRIOR APPLICATION NUMBER: 60/082796
```

RESULT 49
US-09-978-697-614
; Sequence 614, Application US/09978697
; Patent No. US2002016928481
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Baker Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan
; APPLICANT: Ferrara, Napoleon
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Kuo, Sophia S.
; APPLICANT: Napier, Mary A.
; APPLICANT: Pan, James;
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Shelton, David L.
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: P26301P1C27
; CURRENT APPLICATION NUMBER: US/09/978,697
; CURRENT FILING DATE: 2001-10-16

  Db   165 PGAPGPPGPPAEKGAKGAMGRDGATGPSGPGGPPGGVKGEAG------LQGPQGAPG---- 214
  Qy   253 RRRAKPSESRHPGP-----GSFGRGGSLGLGLGGGGMGPREAEPGELGI-------ALRGGG 301
  Db   215 ----KQGATGTPGPQGEKGGSKGDGGLLIGPKGETG----TKGEKGDLGLPGSKGGDRGMKGDA 267
  Qy   302 AADPPFCPHYEKVSGDYGHPVYIVQDGPP 330
  Db   268 GVMGP--PGAQGSKGDFGRP------GPP 288

Query Match   5.2%;   Score 126.5; DB 9; Length 520;
Best Local Similarity   30.9%;   Pred. No. 0.062;
Matches   46;   Conservative   9;   Mismatches   53;   Indels   41;   Gaps   8;

Qy   198 PGKRNLPGDPTSNATS----FGAEGPLPPPSMPAVAGAAGSLALLLGVGAGGAMCWR 252
                                    |||| | | || | ||| ||| || |||

Db
  PRIOR FILING DATE: 1998-04-23
  PRIOR APPLICATION NUMBER: 60/083336
  PRIOR FILING DATE: 1998-04-27
  PRIOR APPLICATION NUMBER: 60/083322
  PRIOR FILING DATE: 1998-04-28
  PRIOR APPLICATION NUMBER: 60/083392
  PRIOR FILING DATE: 1998-04-29
  PRIOR APPLICATION NUMBER: 60/083495
  PRIOR FILING DATE: 1998-04-29
  PRIOR APPLICATION NUMBER: 60/083496
  PRIOR FILING DATE: 1998-04-29
  PRIOR APPLICATION NUMBER: 60/083499
  PRIOR FILING DATE: 1998-04-29
  PRIOR APPLICATION NUMBER: 60/083545
  PRIOR FILING DATE: 1998-04-29
  PRIOR APPLICATION NUMBER: 60/083554
  PRIOR FILING DATE: 1998-04-29
  PRIOR APPLICATION NUMBER: 60/083558
  PRIOR FILING DATE: 1998-04-29
  PRIOR APPLICATION NUMBER: 60/083559
  PRIOR FILING DATE: 1998-04-29
  PRIOR APPLICATION NUMBER: 60/083500
  PRIOR FILING DATE: 1998-04-29
  PRIOR APPLICATION NUMBER: 60/083742
  PRIOR FILING DATE: 1998-04-30
  PRIOR APPLICATION NUMBER: 60/084366
  PRIOR FILING DATE: 1998-05-05
  PRIOR APPLICATION NUMBER: 60/084414
  PRIOR FILING DATE: 1998-05-06
  PRIOR APPLICATION NUMBER: 60/084441
  PRIOR FILING DATE: 1998-05-07
  PRIOR APPLICATION NUMBER: 60/084598
  PRIOR FILING DATE: 1998-05-07
  PRIOR APPLICATION NUMBER: 60/084637
  PRIOR FILING DATE: 1998-05-07
  PRIOR APPLICATION NUMBER: 60/084639
  PRIOR FILING DATE: 1998-05-07
  PRIOR APPLICATION NUMBER: 60/084600
  PRIOR FILING DATE: 1998-05-07
  PRIOR APPLICATION NUMBER: 60/084640
  PRIOR FILING DATE: 1998-05-07
  PRIOR APPLICATION NUMBER: 60/084627
  PRIOR FILING DATE: 1998-05-07
  PRIOR APPLICATION NUMBER: 60/084643
  PRIOR FILING DATE: 1998-05-07
  PRIOR APPLICATION NUMBER: 60/085339
  PRIOR FILING DATE: 1998-05-13
  PRIOR APPLICATION NUMBER: 60/085338
  PRIOR FILING DATE: 1998-05-13
  PRIOR APPLICATION NUMBER: 60/085323
  PRIOR FILING DATE: 1998-05-13
  PRIOR APPLICATION NUMBER: 60/085582
  PRIOR FILING DATE: 1998-05-15
  PRIOR APPLICATION NUMBER: 60/085579
  PRIOR FILING DATE: 1998-05-15
  PRIOR APPLICATION NUMBER: 60/085700
  PRIOR FILING DATE: 1998-05-15
  PRIOR APPLICATION NUMBER: 60/085580
  PRIOR FILING DATE: 1998-05-15
  PRIOR APPLICATION NUMBER: 60/085573
  PRIOR FILING DATE: 1998-05-15
  PRIOR APPLICATION NUMBER: 60/085689
  PRIOR FILING DATE: 1998-05-15
  PRIOR APPLICATION NUMBER: 60/085704
  PRIOR FILING DATE: 1998-05-15
  PRIOR APPLICATION NUMBER: 60/085697

CURRENT FILING DATE: 2001-07-30
  PRIOR APPLICATION NUMBER: 60/062250
  PRIOR FILING DATE: 1997-10-17
  PRIOR APPLICATION NUMBER: 60/064249
  PRIOR FILING DATE: 1997-11-03
  PRIOR APPLICATION NUMBER: 60/065311
  PRIOR FILING DATE: 1997-11-13
  PRIOR APPLICATION NUMBER: 60/066364
  PRIOR FILING DATE: 1997-11-21
  PRIOR APPLICATION NUMBER: 60/077450
  PRIOR FILING DATE: 1998-03-10
  PRIOR APPLICATION NUMBER: 60/077632
  PRIOR FILING DATE: 1998-03-11
  PRIOR APPLICATION NUMBER: 60/077649
  PRIOR FILING DATE: 1998-03-11
  PRIOR APPLICATION NUMBER: 60/077641
  PRIOR FILING DATE: 1998-03-11
  PRIOR APPLICATION NUMBER: 60/077791
  PRIOR FILING DATE: 1998-03-12
  PRIOR APPLICATION NUMBER: 60/078004
  PRIOR FILING DATE: 1998-03-13
  PRIOR APPLICATION NUMBER: 60/078886
  PRIOR FILING DATE: 1998-03-20
  PRIOR APPLICATION NUMBER: 60/078936

~ PRIOR FILING DATE: 1998-03-20
~ PRIOR APPLICATION NUMBER: 60/078910
~ PRIOR FILING DATE: 1998-03-20
~ PRIOR APPLICATION NUMBER: 60/078939
~ PRIOR FILING DATE: 1998-03-20
~ PRIOR APPLICATION NUMBER: 60/079294
~ PRIOR FILING DATE: 1998-03-25
~ PRIOR APPLICATION NUMBER: 60/079656
~ PRIOR FILING DATE: 1998-03-26
~ PRIOR APPLICATION NUMBER: 60/079664
~ PRIOR FILING DATE: 1998-03-27
~ PRIOR APPLICATION NUMBER: 60/079689
~ PRIOR FILING DATE: 1998-03-27
~ PRIOR APPLICATION NUMBER: 60/079663
~ PRIOR FILING DATE: 1998-03-27
~ PRIOR APPLICATION NUMBER: 60/079728
~ PRIOR FILING DATE: 1998-03-27
~ PRIOR APPLICATION NUMBER: 60/079786
~ PRIOR FILING DATE: 1998-03-27
~ PRIOR APPLICATION NUMBER: 60/079920
~ PRIOR FILING DATE: 1998-03-30
~ PRIOR APPLICATION NUMBER: 60/079923
~ PRIOR FILING DATE: 1998-03-30
~ PRIOR APPLICATION NUMBER: 60/080194
~ PRIOR FILING DATE: 1998-03-31
~ PRIOR APPLICATION NUMBER: 60/080165
~ PRIOR FILING DATE: 1998-03-31
~ PRIOR APPLICATION NUMBER: 60/080107
~ PRIOR FILING DATE: 1998-03-31
~ PRIOR APPLICATION NUMBER: 60/080105
~ PRIOR FILING DATE: 1998-04-01
~ PRIOR APPLICATION NUMBER: 60/080328
~ PRIOR FILING DATE: 1998-04-01
~ PRIOR APPLICATION NUMBER: 60/080327
~ PRIOR FILING DATE: 1998-04-01
~ PRIOR APPLICATION NUMBER: 60/080334
~ PRIOR FILING DATE: 1998-04-01
~ PRIOR APPLICATION NUMBER: 60/080333
~ PRIOR FILING DATE: 1998-04-01
~ PRIOR APPLICATION NUMBER: 60/081070
~ PRIOR FILING DATE: 1998-04-08
~ PRIOR APPLICATION NUMBER: 60/081049
~ PRIOR FILING DATE: 1998-04-08
~ PRIOR APPLICATION NUMBER: 60/081071
~ PRIOR FILING DATE: 1998-04-08
~ PRIOR APPLICATION NUMBER: 60/081195
~ PRIOR FILING DATE: 1998-04-08
~ PRIOR APPLICATION NUMBER: 60/081203
~ PRIOR FILING DATE: 1998-04-09
~ PRIOR APPLICATION NUMBER: 60/081229
~ PRIOR FILING DATE: 1998-04-09
~ PRIOR APPLICATION NUMBER: 60/081955
~ PRIOR FILING DATE: 1998-04-15
~ PRIOR APPLICATION NUMBER: 60/081817
~ PRIOR FILING DATE: 1998-04-15
~ PRIOR APPLICATION NUMBER: 60/081819
~ PRIOR FILING DATE: 1998-04-15
~ PRIOR APPLICATION NUMBER: 60/081952
~ PRIOR FILING DATE: 1998-04-15
~ PRIOR APPLICATION NUMBER: 60/081838
~ PRIOR FILING DATE: 1998-04-15
~ PRIOR APPLICATION NUMBER: 60/082568
~ PRIOR FILING DATE: 1998-04-21
~ PRIOR APPLICATION NUMBER: 60/082569
~ PRIOR FILING DATE: 1998-04-21
~ PRIOR APPLICATION NUMBER: 60/082704
~ PRIOR FILING DATE: 1998-04-22
~ PRIOR APPLICATION NUMBER: 60/082804
~ PRIOR FILING DATE: 1998-04-22
~ PRIOR APPLICATION NUMBER: 60/082700
~ PRIOR FILING DATE: 1998-04-22
~ PRIOR APPLICATION NUMBER: 60/082797

~ PRIOR APPLICATION NUMBER: 60/082796
~ PRIOR FILING DATE: 1998-04-23
~ PRIOR APPLICATION NUMBER: 60/083336
~ PRIOR FILING DATE: 1998-04-27
~ PRIOR APPLICATION NUMBER: 60/083322
~ PRIOR FILING DATE: 1998-04-28
~ PRIOR APPLICATION NUMBER: 60/083392
~ PRIOR FILING DATE: 1998-04-29
~ PRIOR APPLICATION NUMBER: 60/083495
~ PRIOR FILING DATE: 1998-04-29
~ PRIOR APPLICATION NUMBER: 60/083496
~ PRIOR FILING DATE: 1998-04-29
~ PRIOR APPLICATION NUMBER: 60/083499
~ PRIOR FILING DATE: 1998-04-29
~ PRIOR APPLICATION NUMBER: 60/083545
~ PRIOR FILING DATE: 1998-04-29
~ PRIOR APPLICATION NUMBER: 60/083554
~ PRIOR FILING DATE: 1998-04-29
~ PRIOR APPLICATION NUMBER: 60/083558
~ PRIOR FILING DATE: 1998-04-29
~ PRIOR APPLICATION NUMBER: 60/083559
~ PRIOR FILING DATE: 1998-04-29
~ PRIOR APPLICATION NUMBER: 60/083500
~ PRIOR FILING DATE: 1998-04-29
~ PRIOR APPLICATION NUMBER: 60/084414
~ PRIOR FILING DATE: 1998-04-30
~ PRIOR APPLICATION NUMBER: 60/083742
~ PRIOR FILING DATE: 1998-04-30
~ PRIOR APPLICATION NUMBER: 60/084366
~ PRIOR FILING DATE: 1998-05-05
~ PRIOR APPLICATION NUMBER: 60/084441
~ PRIOR FILING DATE: 1998-05-06
~ PRIOR APPLICATION NUMBER: 60/084627
~ PRIOR FILING DATE: 1998-05-07
~ PRIOR APPLICATION NUMBER: 60/084637
~ PRIOR FILING DATE: 1998-05-07
~ PRIOR APPLICATION NUMBER: 60/084639
~ PRIOR FILING DATE: 1998-05-07
~ PRIOR APPLICATION NUMBER: 60/084600
~ PRIOR FILING DATE: 1998-05-07
~ PRIOR APPLICATION NUMBER: 60/084598
~ PRIOR FILING DATE: 1998-05-07
~ PRIOR APPLICATION NUMBER: 60/084640
~ PRIOR FILING DATE: 1998-05-07
~ PRIOR APPLICATION NUMBER: 60/084643
~ PRIOR FILING DATE: 1998-05-07
~ PRIOR APPLICATION NUMBER: 60/085339
~ PRIOR FILING DATE: 1998-05-13
~ PRIOR APPLICATION NUMBER: 60/085338
~ PRIOR FILING DATE: 1998-05-13
~ PRIOR APPLICATION NUMBER: 60/085323
~ PRIOR FILING DATE: 1998-05-13
~ PRIOR APPLICATION NUMBER: 60/085582
~ PRIOR FILING DATE: 1998-05-15
~ PRIOR APPLICATION NUMBER: 60/085573
~ PRIOR FILING DATE: 1998-05-15
~ PRIOR APPLICATION NUMBER: 60/085580
~ PRIOR FILING DATE: 1998-05-15
~ PRIOR APPLICATION NUMBER: 60/085579
~ PRIOR FILING DATE: 1998-05-15
~ PRIOR APPLICATION NUMBER: 60/085689
~ PRIOR FILING DATE: 1998-05-15
~ PRIOR APPLICATION NUMBER: 60/085700
~ PRIOR FILING DATE: 1998-05-15
~ PRIOR APPLICATION NUMBER: 60/085704
~ PRIOR FILING DATE: 1998-05-15
~ PRIOR APPLICATION NUMBER: 60/085697

Query Match   5.2%; Score 126.5; DB 9; Length 520;
Best Local Similarity   30.9%; Pred. No. 0.062;
Matches   46; Conservative   9; Mismatches   53; Indels   41; Gaps   8;

Qy   198 PGKENLPGDPTSNATS-----RGAEGPLPPPSMPAVAGAAGGLALLLGVAGAGGAMCWR 252

```
FT   CONFLICT    912    912    P -> R (IN REF. 3).
FT   CONFLICT    933    933    P -> L (IN REF. 3).
FT   CONFLICT    975    975    R -> L (IN REF. 3).
FT   CONFLICT   1064   1064    A -> P (IN REF. 3).
FT   CONFLICT   1084   1084    L -> K (IN REF. 3).
FT   CONFLICT   1120   1120    P -> A (IN REF. 3).
FT   CONFLICT   1123   1123    P -> A (IN REF. 3).
FT   CONFLICT   1126   1126    P -> PPGP (IN REF. 2).
FT   CONFLICT   1206   1206    P -> G (IN REF. 3).
FT   CONFLICT   1206   1206    G -> GQ (IN REF. 3).
FT   CONFLICT   1304   1304    G -> G (IN REF. 3).
FT   CONFLICT   1314   1314    R -> CG (IN REF. 3).
FT   CONFLICT   1323   1324    LR -> T (IN REF. 4).
FT   CONFLICT   1443   1443    A -> G (IN REF. 4).
FT   CONFLICT   1483   1483    S -> Y (IN REF. 4).
FT   SEQUENCE   1516 AA; 153840 MW; 3C70F29A4476EE76 CRC64;

Query Match
Best Local Similarity   5.9%;   Score 144.5; DB 1; Length 1516;
Matches  73;   Conservative 18;   Mismatches 82;   Indels 137;   Gaps 15;

Qy    52  PQIGRDLLCPRARPPGPHSSPNYEFYKLYLVGGQAGRRCEAPPAPNLLLTCDRPDLDL 111
Db   620  PGVGER------------------GPPGPQGPP---------------------------GPPGPS------ 640

Qy   112  RFTIKFQEYSPNLWGHEFRSHHDYYIIATSDGTREG--LESLQGGVCLTRGMKVLLRVGQ 169
Db   641  --FR--HDKLTFIDMEGSGPGGDLEALRG------------------------------ 665

Qy   170  SPRGGAVPRKP--VSEMPMERDRGAAHSLE-PGKENLPGDPTSNATSRGAEGPLPPPSMP 226
Db   666  -PRGPGPPGPPGVPGLPGEPGFGVNSSDVPGPAGLPGVP-------GREGPPGFPGLP 717

Qy   227  AVAGAAGGLALLLLGVAGAGGAMCWRRRAKPSESSRHPGPGSFGRGGSLGLGGGGMGPR 286
Db   718  GPPGPPG----------------REGPPGRTGQKGSLGEAGAPGHKGS 749

Qy   287  EAEPGELGIALRGG-----GAADPPFCP-------HYEKVSGDYGHPVYIV--Q 326
Db   750  KGAPGPAGARGESGLAGAPGPAGPPGPPGPGPLPAGFDDMEGS-GGPFWSTARSA 808

Qy   327  DGPPQGSPPNI 336
Db   809  DG-PQGPPGL 817
```

RESULT 35
```
ID   CA39_CHICK   STANDARD;   PRT;   675 AA.
AC   P32017;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Collagen alpha 3(IX) chain precursor.
GN   COL9A3.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92250566; PubMed=1577778;
RA   Har-El R., Sharma Y.D., Aguilera A., Ueyama N., Wu J.J.,
RA   Eyre D.R., Juricic L., Chandrasekaran S., Li M., Nah H.D.,
RA   Upholt W.B., Tanzer M.L.;
RT   "Cloning and developmental expression of the alpha 3 chain of chicken
RT   type IX collagen.";
RL   J. Biol. Chem. 267:10070-10076(1992).
CC   -!- FUNCTION: COLLAGEN TYPE IX IS A MINOR CARTILAGE NONFIBRILLAR
CC       COLLAGEN. IT IS ASSOCIATED WITH TYPE II COLLAGEN FIBRILS.
CC   -!- SUBUNIT: TRIMERS COMPOSED OF THREE DIFFERENT CHAINS: ALPHA 1(IX),
CC       ALPHA 2(IX), AND ALPHA 3(IX).
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC       UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   used by and for commercial
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M83179; AAB59960.1; -.
DR   PIR; S29218; S22918.
DR   InterPro; IPR000087; Collagen.
DR   Pfam; PF01391; Collagen; 11.
DR   ProDom; PD000007; Collagen; 1.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Glycoprotein; Cartilage; Collagen; Signal.
FT   SIGNAL        1     21
FT   CHAIN        22    675       COLLAGEN ALPHA 3(IX) CHAIN.
FT   DOMAIN       22     25       NONHELICAL REGION 3 (NC3).
FT   DOMAIN       25    515       TRIPLE-HELICAL REGION 3 (COL3).
FT   DOMAIN      516    546       NONHELICAL REGION 2 (NC2).
FT   DOMAIN      547    626       TRIPLE-HELICAL REGION 2 (COL2).
FT   DOMAIN      627    631       NONHELICAL REGION 1 (NC1).
FT   DOMAIN      632    658       TRIPLE-HELICAL REGION 1 (COL1).
FT   DOMAIN      659    675       NONHELICAL REGION 1 (NC1).
FT   SITE        242    244       CELL ATTACHMENT SITE (POTENTIAL).
FT   SITE        591    593       CELL ATTACHMENT SITE (POTENTIAL).
FT   CARBOHYD    479    479       N-LINKED (GLCNAC...) (POTENTIAL).
FT   SEQUENCE    675 AA; 63013 MW; C983FBC924A10098 CRC64;

Query Match
Best Local Similarity   5.9%;   Score 144; DB 1; Length 675;
Matches  64;   Conservative  9;   Mismatches 71;   Indels 78;   Gaps 12;

Qy   166  RVGQSPRGGAVPRKPVSEMPMERDRGAAHSLEPGKENLPGDPTSNATSRGAEGPLPPPSM 225
Db    23  RVG--PQGPPGPRGPPGPSGKDGIDG-------EPGPSGLPGPP-------------GPKGAPGK 65

Qy   226  PAVAGAAG--GLALLLLGVAGAGGAMCWRRRAKPSESSRHPG-------PGSFGRG- 272
Db    66  PGAAGEAGLPGLP-------GVDGLTGT------DGPPGPNGPPGPDRGALGPAGPPGPAGKGL 115

Qy   273  -------GSLGLGGGGGM------GPR-----EAEPGELGIALRGG----------- 300
Db   116  PGPPGPGPSGLPGGNGPFRGPPGPSGLPGFPGPPGPPGPPGLAGIIPEGGGDLQCPALCP 175

RESULT 36
ID   CA19_HUMAN   STANDARD;   PRT;   921 AA.
AC   P20849; Q9Y6P2; Q9Y6P3; Q9H1S1; Q9H1S2; Q99225; Q13699; Q13700;
DT   01-FEB-1991 (Rel. 17, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Collagen alpha 1(IX) chain precursor.
GN   COL9A1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX   MEDLINE=91061164; PubMed=2209617;
RA   Muragaki Y., Kimura T., Ninomiya Y., Olsen B.R.;
RT   "The complete primary structure of two distinct forms of human alpha
RT   1 (IX) collagen chains.";
RL   Eur. J. Biochem. 192:703-708(1990).
RN   [2]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
```

```
RX   MEDLINE=98370844; PubMed=9707347;
RA   Pihlajamaa T., Vuoristo M.M., Annunen S., Perala M., Prockop D.J.,
RA   Ala-Kokko L.;
RT   "Human COL9A1 and COL9A2 genes. Two genes of 90 and 15 kb code for
RT   similar polypeptides of the same collagen molecule.";
RL   Matrix Biol. 17:237-241(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Nickerson T., Harley J.;
RL   Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 580-820 AND 835-884 FROM N.A.
RX   MEDLINE=89137096; PubMed=2465149;
RA   Kimura T., Mattei M.-G., Stevens J.W., Goldring M.B., Ninomiya Y.,
RA   Olsen B.R.;
RT   "Molecular cloning of rat and human type IX collagen cDNA and
RT   localization of the alpha 1(IX) gene on the human chromosome 6.";
RL   Eur. J. Biochem. 179:71-78(1989).
RN   [5]
RP   ALTERNATIVE SPLICING.
RX   MEDLINE=90207204; PubMed=1690886;
RA   Muragaki Y., Nishimura I., Henney A., Ninomiya Y., Olsen B.R.;
RT   "The alpha 1 (IX) collagen gene gives rise to two different
RT   transcripts in both mouse embryonic and human fetal RNA.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:2400-2404(1990).
CC   -!- FUNCTION: Structural component of hyaline cartilage and vitreous
CC       of the eye.
CC   -!- SUBUNIT: Heterotrimer of a alpha 1(IX), a alpha 2(IX) and a alpha
CC       3(IX) chain.
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; A LONG FORM (SHOWN
CC       HERE) AND A SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- DOMAIN: EACH SUBUNIT IS COMPOSED OF THREE TRIPLE-HELICAL DOMAINS
CC       INTERSPERSED WITH NONCOLLAGENOUS DOMAINS. THE GLOBULAR DOMAIN AT
CC       THE N-TERMINUS OF TYPE IX COLLAGEN MOLECULES REPRESENTS THE NC4
CC       DOMAIN WHICH MAY PARTICIPATE IN ELECTROSTATIC INTERACTIONS WITH
CC       POLYANIONIC GLYCOSAMINOGLYCANS IN CARTILAGE.
CC   -!- PTM: Covalently linked to the telopeptides of type II collagen by
CC       lysine-derived cross-links.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC       UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC   -!- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH
CC       INTERRUPTED HELICES (FACIT) FAMILY.
```

```
DR   EMBL; X54412; CAA38276.1; -.
DR   EMBL; X54413; CAA38277.1; -.
DR   EMBL; AF036130; AAC33527.1; JOINED.
DR   EMBL; AF036110; AAC33527.1; JOINED.
DR   EMBL; AF036111; AAC33527.1; JOINED.
DR   EMBL; AF036112; AAC33527.1; JOINED.
DR   EMBL; AF036113; AAC33527.1; JOINED.
DR   EMBL; AF036114; AAC33527.1; JOINED.
DR   EMBL; AF036115; AAC33527.1; JOINED.
DR   EMBL; AF036116; AAC33527.1; JOINED.
DR   EMBL; AF036117; AAC33527.1; JOINED.
DR   EMBL; AF036118; AAC33527.1; JOINED.
DR   EMBL; AF036119; AAC33527.1; JOINED.
DR   EMBL; AF036120; AAC33527.1; JOINED.
DR   EMBL; AF036121; AAC33527.1; JOINED.
DR   EMBL; AF036122; AAC33527.1; JOINED.
DR   EMBL; AF036123; AAC33527.1; JOINED.
DR   EMBL; AF036124; AAC33527.1; JOINED.
DR   EMBL; AF036125; AAC33527.1; JOINED.
DR   EMBL; AF036126; AAC33527.1; JOINED.
DR   EMBL; AF036127; AAC33527.1; JOINED.
DR   EMBL; AF036128; AAC33527.1; JOINED.
DR   EMBL; AF036129; AAC33527.1; JOINED.
DR   EMBL; AF036130; AAC33528.1; -.
DR   EMBL; AF036112; AAC33528.1; JOINED.
DR   EMBL; AF036113; AAC33528.1; JOINED.
DR   EMBL; AF036114; AAC33528.1; JOINED.
DR   EMBL; AF036115; AAC33528.1; JOINED.
DR   EMBL; AF036116; AAC33528.1; JOINED.
DR   EMBL; AF036117; AAC33528.1; JOINED.
DR   EMBL; AF036118; AAC33528.1; JOINED.
DR   EMBL; AF036119; AAC33528.1; JOINED.
DR   EMBL; AF036120; AAC33528.1; JOINED.
DR   EMBL; AF036121; AAC33528.1; JOINED.
DR   EMBL; AF036122; AAC33528.1; JOINED.
DR   EMBL; AF036123; AAC33528.1; JOINED.
DR   EMBL; AF036124; AAC33528.1; JOINED.
DR   EMBL; AF036125; AAC33528.1; JOINED.
DR   EMBL; AF036126; AAC33528.1; JOINED.
DR   EMBL; AF036127; AAC33528.1; JOINED.
DR   EMBL; AF036128; AAC33528.1; JOINED.
DR   EMBL; AF036129; AAC33528.1; JOINED.
DR   EMBL; AL080275; CAC19501.1; ALT_SEQ.
DR   EMBL; AL080275; CAC19502.1; -.
DR   EMBL; AL160262; -; NOT_ANNOTATED_CDS.
DR   EMBL; M32137; AAA53474.1; -.
DR   EMBL; M32133; AAA53474.1; JOINED.
DR   EMBL; M32135; AAA53475.1; -.
DR   EMBL; M32137; AAA53475.1; JOINED.
DR   PIR; S02140; S02140.
DR   PIR; S13580; S13580.
DR   Genew; HGNC;2217; COL9A1.
DR   MIM; 120210; -.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR003129; TSPN.
DR   Pfam; PF01391; Collagen; 30.
DR   Pfam; PF02210; TSPN; 1.
DR   ProDom; PD000007; Collagen; 1.
DR   SMART; SM00210; TSPN; 1.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Glycoprotein; Collagen; Cartilage; Signal; Alternative splicing.
FT   SIGNAL        1     23
FT   CHAIN        24    921       COLLAGEN ALPHA 1(IX) CHAIN.
FT   DOMAIN       24    268       NONHELICAL REGION (NC4).
FT   DOMAIN      269    405       TRIPLE-HELICAL REGION (COL3).
FT   DOMAIN      406    417       NONHELICAL REGION (NC3).
FT   DOMAIN      418    756       TRIPLE-HELICAL REGION (COL2).
FT   DOMAIN      757    786       NONHELICAL REGION (NC2).
FT   DOMAIN      787    901       TRIPLE-HELICAL REGION (COL1).
FT   DOMAIN      902    921       NONHELICAL REGION (NC1).
FT   DISULFID     44    198       BY SIMILARITY.
FT   DISULFID    242    252       BY SIMILARITY.
FT   DISULFID    411    411       INTERCHAIN.
FT   DISULFID    415    415       INTERCHAIN.
FT   CARBOHYD    171    171       N-LINKED (GLCNAC...) (POTENTIAL).
FT   VARSPLIC      1    243       MISSING (IN SHORT ISOFORM).
FT   VARSPLIC    244    267       PLRPRETCHELPARITPSQTTDE -> MAWTARDRGALGL
FT                               LLIGLCLCAAQ (IN SHORT ISOFORM).
FT   CONFLICT    279    280       PP -> AS (IN REF. 1).
FT   CONFLICT    476    476       I -> L (IN REF. 1).
FT   CONFLICT    570    570       Q -> H (IN REF. 2).
FT   CONFLICT    621    621       Q -> R (IN REF. 2).
FT   CONFLICT    910    921       AGQRAFNKGPDP -> LVSEHLTKGLTLERLTAAMLSA
FT                               (IN REF. 1).
SQ   SEQUENCE   921 AA;  91855 MW;  5A6B226B58FC44B34 CRC64;
```

```
Query Match     5.9%;  Score 144;  DB 1;  Length 921;
Best Local Similarity   22.3%;   Pred. No. 0.023;
Matches   84;  Conservative   21;  Mismatches 120;  Indels 152;  Gaps 13;

Qy      2 GPHSGPGGVRVGALLLGVLGVLGVSGLSLEPVWNSANKRFQAEGGYVLYPQIGDRLDLL  61
            |        |||  |     |  | |||            |  |:| || |||||  ||
Db    391 GPP--GPPGPR-------GTIGFHDG---------------------DPL         410
```

RESULT 37

```
Qy   62 CPRAPPPGFHSSPNYEFYKLYLVG------GAQGRRCEAPPAPNLLTCDRPDLDLRFTIK 116
Db  411 CPNACPGRSGYPG----------------LPGMRGHKGAKGEIGE------ 440
Qy  117 FQEYSPNLWGHEFRSHHDYYIIATSDGTREGLSLSLQGGVCLTRGMKVLLRVGQSPRGGAV 176
Db  441 -----PGRQGH---------KGEEGDQG------ELGEVGAQGPPGKAG 468
Qy  177 PRKPVSEMPMERDRGAAHSL--EPGKENLPGDPTSNAT-------SRGAEGPLPP 223
Db  469 GLRGITGIVCDKGEKGARGLDGEPGPQGLPGAPGPQGQGRGPPGEAGPKGDRGARGIP 528
Qy  224 SMPAVAGAAGGLALLLLGVAGAGGAMCWRRRAKPSESRHPG-----PGSFGRGGSLG 276
Db  529 GLPGPKGDTG-----LPGVDGRDGIPGMPGTKGEPGKPGPPGDAGLPGLQGGLPGVPGIPGAKG 583
Qy  277 LGGGGGMGPREAEPGELGTALRGG-----------GAADPPFCPHYEKVSG 316
Db  584 VAGEKGSTGAPGKPGQMGNSGKPGQOGPGEVGPRGPQGLPGSRGELGPVGSPGLPGKLG 643
Qy  317 DYGHPVYIVQDGPPQSP 333
Db  644 SLGSPGLPGLPGPPGLP 660
```

ID   CA1A_CHICK   STANDARD;   PRT;   674 AA.
AC   P08125;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-NOV-1991 (Rel. 20, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Collagen alpha 1(X) chain precursor.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE OF 48-674 FROM N.A. AND SEQUENCE OF 103-117 AND 453-466.
RX   MEDLINE=86168227; PubMed=3082876;
RA   Ninomiya Y., Gordon M., van der Rest M., Schmid T., Linsenmayer T.,
RA   Olsen B.R.;
RT   "The developmentally regulated type X collagen gene contains a long
RT   open reading frame without introns.";
RL   J. Biol. Chem. 261:5041-5050(1986).
RN   [2]
RP   SEQUENCE OF 1-75 FROM N.A.
RX   MEDLINE=89054019; PubMed=2461168;
RA   LuValle P., Ninomiya Y., Rosenblum N.D., Olsen B.R.;
RT   "The type X collagen gene. Intron sequences split the 5'-untranslated
RT   region and separate the coding regions for the non-collagenous amino-
RT   terminal and triple-helical domains.";
RL   J. Biol. Chem. 263:18378-18385(1988).
RN   [3]
RP   REVISIONS TO C-TERMINUS.
RX   MEDLINE=89380199; PubMed=2476437;
RA   Yamaguchi N., Benya P.D., van der Rest M., Ninomiya Y.;
RT   "The cloning and sequencing of alpha 1(VIII) collagen cDNAs
RT   demonstrate that type VIII collagen is a short chain collagen and
RT   contains triple-helical and carboxyl-terminal non-triple-helical
RT   domains similar to those of type X collagen.";
RL   J. Biol. Chem. 264:16022-16029(1989).
CC   -!- FUNCTION: TYPE X COLLAGEN IS A PRODUCT OF HYPERTHROPHIC
CC   CHONDROCYTES AND HAS BEEN LOCALIZED TO PRESUMPTIVE
CC   MINERALIZATION ZONES OF HYALINE CARTILAGE.
CC   -!- SUBUNIT: HOMOTRIMER.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC   UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC   -!- SIMILARITY: STRONG, TO ALPHA 1 AND 2 TYPE VIII COLLAGENS.
CC   -!- SIMILARITY: CONTAINS 1 C1Q DOMAIN.

DR   EMBL; M13496; AAA48736.1; -.
DR   EMBL; J04194; AAA48634.1; ALT_SEQ.
DR   PIR; A31896; A31896.
DR   InterPro; IPR001073; C1q.
DR   InterPro; IPR000087; Collagen.
DR   Pfam; PF00386; C1q; 1.
DR   Pfam; PF01391; Collagen; 8.
DR   PRINTS; PR00007; COMPLEMNTC1Q.
DR   SMART; SM00110; C1Q; 1.
DR   PROSITE; PS01113; C1Q; 1.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Cartilage; Collagen; Signal.
FT   SIGNAL   1   18
FT   CHAIN   19   674   COLLAGEN ALPHA 1 (X) CHAIN.
FT   DOMAIN   19   52   NONHELICAL REGION (NC2)
FT   DOMAIN   53   512   TRIPLE-HELICAL REGION.
FT   DOMAIN   513   674   NONHELICAL REGION (NC1)
FT   DOMAIN   539   674   C1Q.
FT   MOD_RES 453   453   HYDROXYLATION.
FT   MOD_RES 456   456   HYDROXYLATION.
SQ   SEQUENCE   674 AA; 66434 MW; EAB4B1EF174B145 CRC64;

Query Match   5.8%; Score 143; DB 1; Length 674;
Best Local Similarity   25.0%; Pred. No. 0.02;
Matches   77; Conservative 18; Mismatches 91; Indels 122; Gaps 16;

```
Qy   67 PQGPHSSPNYEFY-KLYLVG------GAQGRRCEAPPAPNLLTCDRPDLDLRFTIKF 117
Db   86 PPGPLGPPGFSTVGKLGMPGLPGKPGERGLNGEKGEAGPV----------- 125
Qy  118 QEYSPNLWGHEFRSHHDYYIIATSDGTREGLSLSLQG------GVCLTRGMKVLLRVG-QSP 171
Db  126 --------------GLPGARGFGQGPGIPGPAGLSVLGKPGPGP 156
Qy  172 RGGAVPRKPVSEMPMRDRGAAHSLEPGKENLPGDPTSNAT------------SRGAEG 218
Db  157 PGAQGPRGP----------PGEKGEPGFGVPGINGQKGEMGFGVPGRPGNRGLPG 199
Qy  219 PLPPPSMPAVAGAAGGLALLLLGVAGAGGAMCWRRRAKP---SESRHPGP-GSFGRGGS 274
Db  200 PQGEPGLPGSAGIGKP------GBNGLPGQPGMKGDRGLPGARGEAGIPGPQGGPGSPGE 253
Qy  275 LGLGGGGGMGPREAEPGELGI----ALRG-GGAADPPFCPHYEK--VSGDYGH--PVYIV 325
Db  254 VGIGKPGDMGP---PGPAGIPGAKGLPGPAGLPGSPGLLPGPGFGKPGLLPGMKGHRGP---- 305
Qy  326 QDGPPGSP 333
Db  306 -EGPPGFP 312
```

RESULT 38

ID   CA28_HUMAN   STANDARD;   PRT;   635 AA.
AC   P25067;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Collagen alpha 2(VIII) chain (Endothelial collagen) (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]

RP   SEQUENCE FROM N.A.
RX   MEDLINE=91210292; PubMed=2019595;
RA   Muragaki Y., Jacenko O., Apte S., Mattei M.-G., Ninomiya Y.,
RA   Olsen B.R.;
RT   "The alpha 2(VIII) collagen gene. A novel member of the short chain
RT   collagen family located on the human chromosome 1.";
RL   J. Biol. Chem. 266:7721-7727(1991).
CC   -!- FUNCTION: MAJOR COMPONENT OF THE DESCEMET'S MEMBRANE (BASEMENT
CC   MEMBRANE) OF CORNEAL ENDOTHELIAL CELLS.
CC   -!- SUBUNIT: MAY FORM HOMOTRIMERS, OR HETEROTRIMERS IN ASSOCIATION
CC   WITH ALPHA 1(VIII) TYPE COLLAGENS.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC   UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC   -!- SIMILARITY: STRONG, TO ALPHA 1 TYPES VIII AND X COLLAGENS.
CC   -!- SIMILARITY: CONTAINS 1 C1Q DOMAIN.
CC   ---
CC   ---
DR   EMBL; M60832; AAA62822.1; -.
DR   Genew; HGNC:2216; COL8A2.
DR   MIM; 120252; -.
DR   InterPro; IPR001073; C1q.
DR   InterPro; IPR000087; Collagen.
DR   Pfam; PF00386; C1q; 1.
DR   Pfam; PF01391; Collagen; 8.
DR   SMART; SM00110; C1Q; 1.
DR   PROSITE; PS01113; C1Q; 1.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Cell adhesion; Collagen.
FT   NON_TER   1   1
FT   DOMAIN   1   11   NONHELICAL REGION (NC2).
FT   DOMAIN   12   468   TRIPLE-HELICAL REGION.
FT   DOMAIN   469   635   NONHELICAL REGION (NC1).
FT   DOMAIN   500   635   C1Q.
SQ   SEQUENCE   635 AA; 60527 MW; 22A26116475 4F771 CRC64;

Query Match   5.8%; Score 142.5; DB 1; Length 635;
Best Local Similarity   26.1%; Pred. No. 0.02;
Matches   74; Conservative   17; Mismatches 85; Indels 107; Gaps 14;

Qy 147 GLESLCQ------GVCLTRGMKVLLRVGQSPRGGAVP------RKPVSEMPMERDRGAA 193
              |||||
Db 222 GLPGPQGPSGAKGEPGVTRGPPGCLI----GPTGYGMEGLPGEPGXGDRGPAGVPGLLGDRG-- 275

Qy 194 HSLEPGKENLPGD------PTSNATS------RGAEGPLPPSMPAVAGAAG 233
              |||| |::
Db 276 EPGEDGDGEQGPQGLGGPGPGLPGSAGLPGRGPGPPGLRGEAGPGGPPGVPGIRGDQG 332

Qy 234 --GLALLLLGVAGAGAMCWRRRARKPSESRHPGP------GSFGRG 272
              ||||
Db 333 PSGLA------GKPGVPG------ERGLPGAGPPGPGPGVTGRPGGPPGVAGALGQX 382

Qy 273 GSLGL-------------------GGGGMGPR------EAEPGELGTALRG------GGAAD 304
              ||||
Db 383 GDLGLPGQPGLRGPSGITGPGLLQGPAGPIGPQGLPGLKGEPGLPGPGGEGRAGEPGTAGPRG 442

Qy 305 PPPGPHYEKVSGDYGHPVYIVQDGPPQSPPNIYTSISVLEWP 347
              |:| ||
Db 443 PPGVGPSPGITGPPGLP------GPPGAPGAFDETGIAGLHLP 479

DT   01-APR-1988 (Rel. 07, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
ID   CA25_HUMAN   STANDARD;   PRT;   1496 AA.
AC   P05997;
RESULT 39
CA25_HUMAN

DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Collagen alpha 2 (V) chain precursor.
GN   COL5A2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE OF 1-463 FROM N.A.
RX   MEDLINE=89123368; PubMed=2914927;
RA   Woodbury D., Benson-Chanda V., Ramirez F.;
RT   "Amino-terminal propeptide of human pro-alpha 2(V) collagen conforms
RT   to the structural criteria of a fibrillar procollagen molecule.";
RL   J. Biol. Chem. 264:2735-2738(1989).
RN   [2]
RP   SEQUENCE OF 398-1496 FROM N.A.
RX   MEDLINE=87146331; PubMed=3029669;
RA   Weil D., Bernard M.P., Gargano S., Ramirez F.;
RT   "The pro alpha 2(V) collagen gene is evolutionarily related to the
RT   major fibrillar-forming collagens.";
RL   Nucleic Acids Res. 15:181-198(1987).
RN   [3]
RP   SEQUENCE OF 1227-1496 FROM N.A.
RX   MEDLINE=85289337; PubMed=2411731;
RA   Myers J.C., Loidl H.R., Seyer J.M., Dion A.S.;
RT   "Complete primary structure of the human alpha 2 type V procollagen
RT   COOH-terminal propeptide.";
RL   J. Biol. Chem. 260:11216-11222(1985).
RN   [4]
RP   SEQUENCE OF 1449-1496 FROM N.A.
RX   MEDLINE=89138450; PubMed=3224983;
RA   Tsipouras P., Schwartz R.C., Liddell A.C., Salkeld C.S., Weil D.,
RA   Ramirez F.;
RT   "Genetic distance of two fibrillar collagen loci, COL3A1 and COL5A2,
RT   located on the long arm of human chromosome 2.";
RL   Genomics 3:275-277(1988).
RN   [5]
RP   SEQUENCE OF 208-227.
RC   TISSUE=Placenta;
RX   MEDLINE=92239022; PubMed=1571108;
RA   Mann K.;
RT   "Isolation of the alpha 3-chain of human type V collagen and
RT   characterization by partial sequencing.";
RL   Biol. Chem. Hoppe-Seyler 373:69-75(1992).
RN   [6]
RP   SEQUENCE OF 288-297 AND 606-617.
RC   TISSUE=Bone;
RX   MEDLINE=94237164; PubMed=8181482;
RA   Moradi-Ameli M., Rousseau J.C., Kleman J.P., Champliaud M.F.,
RA   Boutillon M.M., Bernillon J., Wallach J.M., van der Rest M.;
RT   "Diversity in the processing events at the N-terminus of type-V
RT   collagen.";
RL   Eur. J. Biochem. 221:987-995(1994).
RN   [7]
RP   DISEASE.
RX   PubMed=9425231;
RA   Michalickova K., Susic M., Willing M.C., Wenstrup R.J., Cole W.G.;
RT   "Mutations of the alpha2(V) chain of type V collagen impair matrix
RT   assembly and produce Ehlers-Danlos syndrome type I.";
RL   Hum. Mol. Genet. 7:249-255(1998).
RN   [8]
RP   VARIANT EDS-II ARG-960.
RX   PubMed=9783710;
RA   Richards A.J., Martin S., Nicholls A.C., Harrison J.B., Pope F.M.,
RA   Burrows N.P.;
RT   "A single base mutation in COL5A2 causes Ehlers-Danlos syndrome type
RT   II.";
RL   J. Med. Genet. 35:846-848(1998).
CC   -!- FUNCTION: TYPE V COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
CC   (FIBRILLAR FORMING COLLAGEN). IT IS A MINOR CONNECTIVE TISSUE
CC   COMPONENT OF NEARLY UBIQUITOUS DISTRIBUTION. TYPE V COLLAGEN BINDS
CC   TO DNA, HEPARAN SULFATE, THROMBOSPONDIN, HEPARIN, AND INSULIN.
CC   -!- SUBUNIT: TRIMERS OF TWO ALPHA 1(V) AND ONE ALPHA 2(V) CHAINS IN

Left column (COL5A2 entry, continued)

```
CC       MOST TISSUES AND TRIMERS OF ONE ALPHA 1(V), ONE ALPHA 2(V), AND
CC       ONE ALPHA 3(V) CHAINS IN PLACENTA.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC       UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC   -!- DISEASE: Defects in COL5A2 are a cause of Ehlers-Danlos syndrome
CC       gravis or type I (EDS-I), and Ehlers-Danlos syndrome mitis or type
CC       II (EDS-II). Ehlers-Danlos syndrome is a genetically and
CC       phenotypically heterogeneous connective-tissue disorder
CC       characterized by loose-jointedness and fragile, velvety and
CC       stretchable, bruisable skin that heals with peculiar 'cigarette-
CC       paper' scars. EDS-I and EDS-II are autosomal dominant traits.
CC   -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
```

```
DR   EMBL; J04478; AAA51859.1; -.
DR   EMBL; X04758; CAA28454.1; -.
DR   EMBL; M11718; AAA52058.1; -.
DR   PIR; A25374; A25374.
DR   PIR; A25874; A25874.
DR   PIR; A30017; A30017.
DR   PIR; A31427; A31427.
DR   Genew; HGNC:2210; COL5A2.
DR   MIM; 120190; -.
DR   MIM; 130000; -.
DR   MIM; 130010; -.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001007; VWF_C.
DR   Pfam; PF00093; vwc; 1.
DR   Pfam; PF01391; Collagen; 18.
DR   Pfam; PF01410; COLFI; 1.
DR   ProDom; PD000007; Collagen; 5.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS01208; VWFC; 1.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Glycoprotein; Collagen; Signal; Ehlers-Danlos syndrome;
KW   Disease mutation.
FT   SIGNAL        1     26
FT   CHAIN        27   1226       COLLAGEN ALPHA 2(V) CHAIN.
FT   PROPEP     1227   1496       CARBOXYL-TERMINAL PROPEPTIDE.
FT   DOMAIN       39     97       VWFC.
FT   MOD_RES     290    290       HYDROXYLATION.
FT   MOD_RES     293    293       HYDROXYLATION.
FT   MOD_RES     296    296       HYDROXYLATION.
FT   MOD_RES     608    608       HYDROXYLATION.
FT   MOD_RES     614    614       HYDROXYLATION.
FT   VARIANT     960    960       G -> R (IN EDS-II).
FT                                /FTId=VAR_013588.
FT   CONFLICT    292    292       A -> P (IN REF. 6).
FT   CONFLICT   1418   1418       A -> T (IN REF. 3).
FT   CONFLICT   1418   1418       K -> T (IN REF. 3).
FT   CONFLICT   1438   1438       F -> S (IN REF. 3).
FT   CONFLICT   1460   1460       E -> Q (IN REF. 4).
FT   CONFLICT   1496   1496       V -> A (IN REF. 4).
SQ   SEQUENCE   1496 AA;  144720 MW;  82827C17A8644F5A CRC64;
```

```
Query Match     5.8%;  Score 142.5;  DB 1;  Length 1496;
Best Local Similarity   29.1%;  Pred. No. 0.048;
Matches    55;  Conservative   15;  Mismatches   76;  Indels   43;  Gaps    9;

Qy   169 QSPRG--GAV--PRKPVSEMPMERDRGAAHSLEPGKENLPG--------DPTSNATSRGAE 217
             ||||    |   ||| |                            |  || | | |||
Db   500 RGPRGDPGTLGPPGPVGERGAPGNRGF-----PGSDGLPGPRGAQGERGPVGSSGPKGSQ 554

Qy   218 GPLPPPSMPAVAGAAGGLALILLGVAGAGGAMCWRRRAKPSESRHPGP---GSFGRGG 273
```

Right column (RESULT 40)

```
Db   555 GDPGPGEPGLPGPARG-----LTGNPGVGGPEGPLGPDGAPGEDGRPGRPGGSIGKGQPG 609
Qy   274 SLGLGGGGMGPREAPFGELGIALRGGGAADPFCFHYEKVSGDYGHPVYIVQDGPP---   330
Db   610 TMGLPGPKGSNGDPGKPGEAG-----NPGVPGQRGAPGPDGKVGPYGPP------GPGLR 659
Qy   331 -----QSPP 334
Db   660 GERGSQGPP 668
```

```
RESULT 40

ID   CA18_HUMAN     STANDARD;      PRT;    744 AA.
AC   P27658; Q96D07;
DT   01-AUG-1992 (Rel. 23, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Collagen alpha 1(VIII) chain precursor (Endothelial collagen).
GN   COL8A1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lung;
RA   Strausberg R.;
RL   Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91231001; PubMed=2029894;
RA   Muragaki Y., Mattei M.-G., Yamaguchi N., Olsen B.R., Ninomiya Y.;
RT   "The complete primary structure of the human alpha 1 (VIII) chain and
RT   assignment of its gene (COL8A1) to chromosome 3.";
RL   Eur. J. Biochem. 197:615-622(1991).
CC   -!- FUNCTION: MAJOR COMPONENT OF THE DESCEMET'S MEMBRANE (BASEMENT
CC       MEMBRANE) OF CORNEAL ENDOTHELIAL CELLS.
CC   -!- SUBUNIT: MAY FORM HOMOTRIMERS, OR HETEROTRIMERS IN ASSOCIATION
CC       WITH ALPHA 2 (VIII) TYPE COLLAGENS.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC       UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC   -!- MISCELLANEOUS: FOUR CONSECUTIVE GLY-PRO-PRO TRIPLETS ARE PRESENT
CC       AT THE C-TERMINUS OF THE TRIPLE-HELICAL REGION. THESE MAY PROVIDE
CC       THE HIGH THERMAL STABILITY OF THIS REGION.
CC   -!- SIMILARITY: STRONG, TO ALPHA 2 TYPES VIII AND X COLLAGENS.
CC   -!- SIMILARITY: CONTAINS 1 C1Q DOMAIN.
```

```
DR   EMBL; X57527; CAA40748.1; -.
DR   EMBL; BC013581; AAH13581.1; -.
DR   PIR; S15435; S15435.
DR   Genew; HGNC:2215; COL8A1.
DR   MIM; 120251; -.
DR   InterPro; IPR001073; C1q.
DR   InterPro; IPR000087; Collagen.
DR   Pfam; PF00386; C1q; 1.
DR   Pfam; PF01391; Collagen; 8.
DR   PRINTS; PR0007; COMPLEMENTC1Q.
DR   ProDom; PD000007; Collagen; 1.
DR   SMART; SM00110; C1Q; 1.
DR   PROSITE; PS01113; C1Q; 1.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Glycoprotein; Cell adhesion; Collagen; Signal.
FT   SIGNAL        1     28
```

```
FT   CHAIN         29    744       COLLAGEN ALPHA 1 (VIII) CHAIN.
FT   DOMAIN        29    117       NONHELICAL REGION (NC2).
FT   DOMAIN       118    571       TRIPLE-HELICAL REGION (COL1).
FT   DOMAIN       572    744       NONHELICAL REGION (NC1).
FT   DOMAIN       609    744       C1Q.
FT   CONFLICT     262    262       P -> L (IN REF. 1).
FT   CONFLICT     297    297       P -> R (IN REF. 1).
FT   CONFLICT     344    344       P -> A (IN REF. 1).
FT   CONFLICT     382    382       P -> S (IN REF. 1).
FT   CONFLICT     388    388       A -> L (IN REF. 1).
FT   CONFLICT     454    454       L -> F (IN REF. 1).
FT   CONFLICT     464    464       A -> H (IN REF. 1).
FT   CONFLICT     601    601       Y -> T (IN REF. 1).
FT   CONFLICT     631    631       A -> G (IN REF. 1).
FT   SEQUENCE     744 AA;  73364 MW;  2BC1B0955DE2C9A3 CRC64;
```

Query Match   5.8%; Score 142; DB 1; Length 744;
Best Local Similarity   23.3%; Pred. No. 0.025;
Matches   72; Conservative   16; Mismatches   99; Indels 122; Gaps   13;

```
Qy      67 PPGPHSSPNYEFYKLYLVGGAQGRRCBAPPAPNLLLTCDRPDLDLRFTIKPQEYSPNLWG 126
                                                                           
Db     198 PPGPHGLPG-----------------IGKPGPGLPGPGPK 222

Qy     127 HEFRSHHDYYIIATSDGTREGLESLGGGVCLITRGMVKVLLRVGQSPRGGAVPRKPVSEMPM 186
                                                                           
Db     223 --------------GDRGPKGLPGPG------------ 256

Qy     187 ERDRGAAHSLEPGKENLPGDPTSNATS-RGAEGPLPPPSMPAVAGAAGGLALLLL----G 241
                          |||         |  |||                               
Db     257 ----PGMHG-PPGPVGLPGVGKPGVTGPPGPGPGPLGKPGAPGEPGPQGPPIGVPGVQGPPG 311

Qy     242 VAGAGGAMCMRRRRAKPSESRHPG--------------------PGSFGRGGS 274
                                                                           
Db     312 IPGIG------------KPGQODIPGPGPGFPGPGGKGBOGLPGLPGPGGLPGIGKPGFPGPGKGD 361

Qy     275 LGLGG-GGGMGPREABPGELGIALRGGGAADPPFCPHYEKVSGDYGHPVYIVQDGP---- 329

Qy     330 ----PGSPP 334

Db     362 RGMGGVPGALGPR-GEKGPIGAPGIGG----PPGEPGLPGIPGPMGPPGAIGFPGPKGEG 416

Db     417 GIVGPQGPP 425
```

RESULT 41
CA26_MOUSE
```
ID   CA26_MOUSE     STANDARD;      PRT;  1029 AA.
AC   Q02778; Q05505;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Collagen alpha 2(VI) chain precursor.
GN   COL6A2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ibrahimi A., Bardon S., Dani C.;
RL   Submitted (MAY-1992) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 266-1029 FROM N.A.
RX   MEDLINE=93143659; PubMed=8380980;
RA   Ibrahimi A., Bertrand B., Bardon S., Amri E.Z., Grimaldi P.,
RA   Ailhaud G., Dani C.;
RT   "Cloning of alpha 2 chain of type VI collagen and expression during
RT   mouse development.";
RL   Biochem. J. 289:141-147(1993).
RN   [3]
RP   SEQUENCE OF 266-600 FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Fibroblast;
RX   MEDLINE=91226374; PubMed=1709252;
RA   Constantinou C.D., Jimenez S.A.;
RT   "Structure of cDNAs encoding the triple-helical domain of murine
RT   alpha 2 (VI) collagen chain and comparison to human and chick
RT   homologues. Use of polymerase chain reaction and partially degenerate
RT   oligonucleotide for generation of novel cDNA clones.";
RL   Matrix 11:1-9(1991).
RN   [4]
RP   SEQUENCE OF 659-1029 FROM N.A.
RX   MEDLINE=93256888; PubMed=8493506;
RA   Zhang R.Z., Pan T.C., Timpl R., Chu M.-L.;
RT   "Cloning and sequence analysis of cDNAs encoding the alpha 1, alpha 2
RT   and alpha 3 chains of mouse collagen VI.";
RL   Biochem. J. 291:787-792(1993).
CC   -!- FUNCTION: COLLAGEN VI ACTS AS A CELL-BINDING PROTEIN.
CC   -!- SUBUNIT: TRIMERS COMPOSED OF THREE DIFFERENT CHAINS: ALPHA 1(VI),
CC       ALPHA 2(VI), AND ALPHA 3(VI).
CC   -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN ADIPOSE TISSUE, LUNG,
CC       KIDNEY, HEART, INTESTINE AND OVARY. LOWER LEVELS IN TESTIS, TONGUE, SKIN,
CC       ADRENAL GLANDS AND OVARY. NO EXPRESSION IN SKELETAL
CC       MUSCLE OR LIVER.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC       UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC   -!- SIMILARITY: CONTAINS 3 VWFA DOMAINS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X65582; CAA46541.1; -.
DR   EMBL; X62332; CAA44206.1; -.
DR   EMBL; X62332; CAA44206.1; -.
DR   EMBL; Z18272; CAA79153.1; -.
DR   EMBL; L06343; AAA37441.1; -.
DR   PIR; S13745; S13745.
DR   MGD; MGI:88460; Col6a2.
DR   InterPro; IPR002035; Collagen.
DR   InterPro; IPR000087; Collagen.
DR   Pfam; PF00092; vwa; 3.
DR   Pfam; PF01391; Collagen; 4.
DR   PRINTS; PR00453; VWFADOMAIN.
DR   SMART; SM00327; VWA; 3.
DR   PROSITE; PS50234; VWFA; 3.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Glycoprotein; Cell adhesion; Collagen; Signal.
FT   SIGNAL         1     20
FT   CHAIN         21   1029       COLLAGEN ALPHA 2 (VI) CHAIN.
FT   DOMAIN        21    265       NONHELICAL REGION.
FT   DOMAIN       266    600       TRIPLE-HELICAL REGION.
FT   DOMAIN       601   1029       NONHELICAL REGION.
FT   DOMAIN        54    178       VWFA 1.
FT   DOMAIN       623    748       VWFA 2.
FT   DOMAIN       843    965       VWFA 3.
FT   CARBOHYD     150    150       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     337    337       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     640    640       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     795    795       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     907    907       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     963    963       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CONFLICT     268    268       H -> S (IN REF. 3).
FT   CONFLICT     804    804       A -> S (IN REF. 4).
FT   CONFLICT     848    848       Q -> L (IN REF. 4).
FT   CONFLICT     962    966       TGNDS -> (IN REF. 4).
FT   CONFLICT     976    977       TR -> KQ (IN REF. 4).
FT   SEQUENCE    1029 AA;  109811 MW;  EDA1BA0AF266FF5D CRC64;
```

RESULT 42

Query Match   Best Local Similarity   29.7%;   Score 142;   DB 1;   Length 1029;
                                        Pred. No. 0.035;
Matches   90;   Conservative   16;   Mismatches 125;   Indels 72;   Gaps 20;

```
Qy    68  PGPHSSPNYEFYKILVLVGGAQGRRCEAAPAPNLLLTCDRPDLLDLRFTIKFQEYSPNLWGH  127
          |||||   | |||||
Db   265  PGPHGPKGYRGQK------GAKGNMGE-PGEPGQKGRQGDPGIGEPGIPGPGKGVPGFKGE   318

Qy   128  --EPRSHHDYYIIATSDGTREGLESLQG--GVCLTRGMKVLL-RVG-QSPRGGAVPRKP-   180

Db   319  KGEF-----------GSDG-RKGAPGLAGKNG---TDGQKGKLGRIGPPGCKGDPGSRGPD  364

Qy   181  ---VSEMPMER-DRGA-AHSLEPGKENLPGDP?SNATSRGAEGPLPPPSMPAVAGAAGG    234

Db   365  GYPGEAGSPGERGDQGAKGDSGRPGRRGPPGDP-GDKGSKGYQGNNGAPGSPGVKGGKGKG  423

Qy   235  LALLLGVAGAGANCWRRRAKP----------SESRHPGP----GSFGRGGSLGLG       278

Db   424  -----PGPRGPKGEPGRRGDPGTKGGPGPGSDGPPGKGDPGPGEGPRGLAGEVGSKGAK   475

Qy   279  GGGGM-GPREAEPELGIAIRGGGAADPPFCPHYEKVSGDYGHPVYIVQDGPPQSP- PN   335

Db   476  GDRGLPGPRGPQ-GALGEPGKQGSRGDP-------------GDAGPRGDSGQPGPRKGDPGRRG  524

Qy   336  IYY   338

Db   525  FSY   527
```

ID   CA24_ASCSU   STANDARD;   PRT;   1763 AA.
AC   P27393;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Collagen alpha 2(IV) chain precursor.
OS   Ascaris suum (Pig roundworm) (Ascaris lumbricoides).
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea;
OC   Ascarididae; Ascaris.
OX   NCBI_TaxID=6253;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93340768; PubMed=1714907;
RA   Pettitt J., Kingston I.B.;
RT   "The complete primary structure of a nematode alpha 2(IV) collagen
RT   and the partial structural organization of its gene.";
RL   J. Biol. Chem. 266:16149-16156(1991).
CC   -!- FUNCTION: COLLAGEN TYPE IV IS SPECIFIC FOR BASEMENT MEMBRANES.
CC   -!- SUBUNIT: TRIMERS OF TWO ALPHA 1(IV) AND ONE ALPHA 2(IV) CHAIN.
CC   TYPE IV COLLAGEN FORMS A MESH-LIKE NETWORK LINKED THROUGH
CC   INTERMOLECULAR INTERACTIONS BETWEEN 7S DOMAINS AND BETWEEN NC1
CC   DOMAINS.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; I (SHOWN HERE) AND II; ARE
CC   PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
CC   DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
CC   G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
CC   CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
CC   TRIPLE-HELICAL 7S DOMAIN.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC   UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC   -!- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTINE RESIDUES WHICH
CC   ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
CC   THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
CC   IV COLLAGENS.

DR   EMBL; M67507; AAA18014.1; -.
DR   PIR; S16366; S16366.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR001442; ProcollagnC4.
DR   Pfam; PF01391; Collagen; 25.
DR   Pfam; PF01413; C4; 2.
DR   ProDom; PD000007; Collagen; 3.
DR   ProDom; PD003923; ProcollagnC4; 2.
DR   SMART; SM00111; C4; 2.
KW   Collagen; Repeat; Collagen;
KW   Alternative splicing; Connective tissue; Basement membrane; Repeat; Collagen;
KW   Hydroxylation; Glycoprotein; Signal.
FT   SIGNAL   1   26   POTENTIAL.
FT   CHAIN   27   1763   COLLAGEN ALPHA 2 (IV) CHAIN.
FT   DOMAIN   27   42   7S DOMAIN.
FT   DOMAIN   43   1529   TRIPLE-HELICAL REGION.
FT   DOMAIN   1530   1763   NONHELICAL REGION (NC1).
FT   DISULFID   1548   1637   (BY SIMILARITY).
FT   DISULFID   1581   1634   (BY SIMILARITY).
FT   DISULFID   1593   1599   (BY SIMILARITY).
FT   DISULFID   1656   1752   (BY SIMILARITY).
FT   DISULFID   1690   1749 OR 1749 (BY SIMILARITY).
FT   DISULFID   1702   1752 OR 1752 (BY SIMILARITY).
FT   DISULFID   1709   BY SIMILARITY.
FT   CARBOHYD   126   126   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   249   249   O-LINKED (GLYCOSAMINOGLYCAN) (IN FORM II)
FT   (POTENTIAL).
FT   VARSPLIC   230   266   GEQGPRGPQGPGPVPSTGAKGTIIGPEBGAPGMKGEK ->
FT   GDIGPAGPPGPPGPREFTGSGSIVGPRGHSGDKGVK (IN
FT   ISOFORM II).
SQ   SEQUENCE   1763 AA;   304F5288C06AAE0D CRC64;

RESULT 43

Query Match   Best Local Similarity   25.5%;   Score 142;   DB 1;   Length 1763;
                                        Pred. No. 0.061;
Matches   97;   Conservative   24;   Mismatches 127;   Indels 132;   Gaps 23;

```
Qy     2  GPPHSGPGGVRVGALLLLGVLGL-----VSGLSLEP------VYWNSANKRFQAEGGVVLY    51

Db    82  GPP--GPQGIK-GDRGIIGVPGFPGNDGANGRPGEGPPGAPGWDGCNGTDGAPG---V    134

Qy    52  PQIGDRLDLLCPRARP--PGPHSSPNYEFYKILVLVGGAQGRRCEA--PPAPNLLLTCDRP   107

Db   135  P-----------GLPGPGMGPGPPGPVPGMKGEPAIGYAGAPGEKGDAGMPGMPGL------  181

Qy   108  DLDLRFTIKFQEYSPNLWGHEFRSHHDYYIIATSDGTREGLESLQGGVCLTRGMKVLLRV   167

Db   182  ----------PGPPG------------RDGFPGEKG------DRG------DV        200

Qy   168  GQS-PRG------------------GAV------PRKFVSEMPMERDRGAAHSLE--P    198

Db   201  GQAGPRGPGPPGEAGPPGENPGIGIGPKGDPGDPGEQGPRGPQGPGPGPVPSTGAKGTIIGPEGAP  260

Qy   199  GKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAGGLALLLLGVAGAGANCWRRRAKP     258

Db   261  GMKGEKGDP--------GEAGPRGFPGTPGVAQGPG-----LPGMKGEKGLSGPAGPRGKE   308

Qy   259  SESRHPGGSFGRGGSLGLGGGGGMGPREAEPELGIALRGG--GAADPPFCPHYEKVSG     316

Db   309  GRPGLPGPPGF--KGDRGLDLPGVPGLPGPGKGEAGPPGRDGAKGRARGPPGPPG---GG   362

Qy   317  DYGHPVIVQDGPPQSPPNI   336

Db   363  EF------SDGPP-GPPGL   374
```

ID   CA13_HUMAN   STANDARD;   PRT;   1466 AA.
AC   P02461; Q15112;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)

```
DE   Collagen alpha 1 (III) chain precursor.
GN   COL3A1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Skin fibroblast;
RX   MEDLINE=89350838; PubMed=2764886;
RA   Ala-Kokko L., Kontusaari S., Baldwin C.T., Kuivaniemi H.,
RA   Prockop D.J.;
RT   "Structure of cDNA clones coding for the entire prepro alpha 1 (III)
RT   chain of human type III procollagen. Differences in protein structure
RT   from type I procollagen and conservation of codon preferences.";
RL   Biochem. J. 260:509-516(1989).
RN   [2]
RP   SEQUENCE OF 149-1225 FROM N.A.
RX   MEDLINE=89386015; PubMed=2780304;
RA   Janeczko R.A., Ramirez F.;
RT   "Nucleotide and amino acid sequences of the entire human alpha 1
RT   (III) collagen.";
RL   Nucleic Acids Res. 17:6742-6742(1989).
RN   [3]
RP   SEQUENCE OF 168-398.
RX   MEDLINE=77134724; PubMed=557335;
RA   Seyer J.M., Kang A.H.;
RT   "Covalent structure of collagen: amino acid sequence of cyanogen
RT   bromide peptides from the amino-terminal segment of type III collagen
RT   of human liver.";
RL   Biochemistry 16:1158-1164(1977).
RN   [4]
RP   SEQUENCE OF 168-398.
RX   MEDLINE=79000343; PubMed=687591;
RA   Seyer J.M., Kang A.H.;
RT   "Covalent structure of collagen: amino acid sequence of five
RT   consecutive CNBr peptides from type III collagen of human liver.";
RL   Biochemistry 17:3404-3411(1978).
RN   [5]
RP   SEQUENCE OF 399-727.
RA   Seyer J.M.;
RL   Submitted (DEC-1977) to the PIR data bank.
     REVISIONS.
RN   [6]
RP   SEQUENCE OF 728-964.
RX   MEDLINE=80198282; PubMed=6246925;
RA   Seyer J.M., Mainardi C., Kang A.H.;
RT   "Covalent structure of collagen: amino acid sequence of alpha 1
RT   (III)-CB5 from type III collagen of human liver.";
RL   Biochemistry 19:1583-1589(1980).
RN   [7]
RP   SEQUENCE OF 950-1466 FROM N.A.
RX   MEDLINE=88189827; PubMed=3357782;
RA   Mankoo B.S., Dalgleish R.;
RT   "Human pro alpha 1 (III) collagen: cDNA sequence for the 3' end.";
RL   Nucleic Acids Res. 16:2337-2337(1988).
RN   [8]
RP   SEQUENCE OF 1065-1466 FROM N.A.
RX   MEDLINE=85087944; PubMed=6096827;
RA   Loidl H.R., Brinker J.M., May M., Pihlajaniemi T., Morrow S.,
RA   Rosenbloom J., Myers J.C.;
RT   "Molecular cloning and carboxyl-propeptide analysis of human type III
RT   procollagen.";
RL   Nucleic Acids Res. 12:9383-9394(1984).
RN   [9]
RP   REVISION TO 1184.
RA   Molyneux K., Dalgleish R.;
RT   "Human type III collagen 'variant' is a cDNA cloning artefact.";
RL   Nucleic Acids Res. 16:11833-11833(1988).
RN   [10]
RP   SEQUENCE OF 965-1200.
RX   MEDLINE=81206139; PubMed=7016180;
RA   Seyer J.M., Kang A.H.;
RT   "Covalent structure of collagen: amino acid sequence of alpha 1
RT   (III)-CB9 from type III collagen of human liver.";
RL   Biochemistry 20:2621-2627(1981).
RN   [11]
RP   SEQUENCE OF 1176-1466 FROM N.A.
RX   MEDLINE=85157600; PubMed=2579949;
RA   Chu M.-L., Weil D., de Wet W.J., Bernard M.P., Sippola M., Ramirez F.;
RT   "Isolation of cDNA and genomic clones encoding human pro-alpha 1
RT   (III) collagen. Partial characterization of the 3' end region of the
RT   gene.";
RL   J. Biol. Chem. 260:4357-4363(1985).
RN   [12]
RP   SEQUENCE OF 1161-1200 FROM N.A.
RX   MEDLINE=86187804; PubMed=3754462;
RA   Miskulin M., Dalgleish R., Kluve-Beckerman B., Rennard S.I.,
RA   Tolstoshev P., Brantly M., Crystal R.G.;
RT   "Human type III collagen gene expression is coordinately modulated
RT   with the type I collagen genes during fibroblast growth.";
RL   Biochemistry 25:1408-1413(1986).
RN   [13]
RP   SEQUENCE OF 1-170 FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=88303360; PubMed=3405773;
RA   Toman D., Ricca G., de Crombrugghe B.;
RT   "Nucleotide sequence of a cDNA coding for the amino-terminal region
RT   of human prepro alpha 1(III) collagen.";
RL   Nucleic Acids Res. 16:7201-7201(1988).
RN   [14]
RP   SEQUENCE OF 1-176 FROM N.A.
RX   MEDLINE=89378752; PubMed=2777083;
RA   Benson-Chanda V., Su M.W., Weil D., Chu M.-L., Ramirez F.;
RT   "Cloning and analysis of the 5' portion of the human type-III
RT   procollagen gene (COL3A1).";
RL   Gene 78:255-265(1989).
RN   [15]
RP   REVIEW ON VARIANTS.
RX   MEDLINE=97255959; PubMed=9101290;
RA   Kuivaniemi H., Tromp G., Prockop D.J.;
RT   "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
RT   associated collagen (type IX), and network-forming collagen (type X)
RT   cause a spectrum of diseases of bone, cartilage, and blood vessels.";
RL   Hum. Mutat. 9:300-315(1997).
RN   [16]
RP   VARIANT AORTIC ANEURYSM ARG-303, AND VARIANT THR-668.
RX   MEDLINE=93293988; PubMed=8514866;
RA   Tromp G., Wu Y., Prockop D.J., Madhatheri S.L., Kleinert C.,
RA   Earley J.J., Zhang J., Noergaard O., Darling R.C., Abbott W.M.,
RA   Cole C.W., Jaakkola P., Ryynaenen M., Pearce W.H., Yao J.S.T.,
RA   Majamaa K., Smullens S.N., Gatalica Z., Ferrell R.E., Jimenez S.A.,
RA   Jackson C.E., Michels V.V., Kaye M., Kuivaniemi H.;
RT   "Sequencing of cDNA from 50 unrelated patients reveals that mutations
RT   in the triple-helical domain of type III procollagen are an
RT   infrequent cause of aortic aneurysms.";
RL   J. Clin. Invest. 91:2539-2545(1993).
RN   [17]
RP   VARIANT THR-698.
RX   MEDLINE=91045136; PubMed=2235526;
RA   Zafarullah K., Kleinert C., Tromp G., Kuivaniemi H., Kontusaari S.,
RA   Wu Y., Ganguly A., Prockop D.J.;
RT   "G to A polymorphism in exon 31 of the COL3A1 gene.";
RL   Nucleic Acids Res. 18:6180-6180(1990).
RN   [18]
RP   VARIANT AORTIC ANEURYSM ARG-786.
RX   MEDLINE=91056145; PubMed=2243125;
RA   Kontusaari S., Tromp G., Kuivaniemi H., Romanic A.M., Prockop D.J.;
RT   "A mutation in the gene for type III procollagen (COL3A1) in a family
RT   with aortic aneurysms.";
RL   J. Clin. Invest. 86:1465-1473(1990).
RN   [19]
RP   VARIANT EDS-IV ARG-828.
RX   MEDLINE=94016385; PubMed=8411057;
RA   Richards A.J., Narcisi P., Lloyd J.C., Ferguson C., Pope F.M.;
RT   "The substitution of glycine 661 by arginine in type III collagen
```

RESULT 44
CA18_RABIT

RT produces mutant molecules with different thermal stabilities and
RT causes Ehlers-Danlos syndrome type IV.";
RL J. Med. Genet. 30:690-693(1993).
RN [20]
RP VARIANT EDS-IV SER-957.
RX MEDLINE=89109135; PubMed=2492273;
RA Tromp G., Kuivaniemi H., Shikata H., Prockop D.J.;
RT "A single base mutation that substitutes serine for glycine 790 of
RT the alpha 1 (III) chain of type III procollagen exposes an arginine
RT and causes Ehlers-Danlos syndrome type IV.";
RL J. Biol. Chem. 264:1349-1352(1989).
RN [21]
RP VARIANT EDS-IV GLU-1014.
RX MEDLINE=92316511; PubMed=1352273;
RA Richards A.J., Ward P.N., Narcisi P., Nicholls A.C., Lloyd J.C.,
RA Pope F.M.;
RT "A single base mutation in the gene for type III collagen (COL3A1)
RT converts glycine 847 to glutamic acid in a family with Ehlers-Danlos
RT syndrome type IV. An unaffected family member is mosaic for the
RT mutation.";
RL Hum. Genet. 89:414-418(1992).
RN [22]
RP VARIANT EDS-IV VAL-960.
RX MEDLINE=9526429; PubMed=7749417;
RA Tromp G., de Paepe A., Nuytinck L., Madhatheri S.L., Kuivaniemi H.;
RT "Substitution of valine for glycine 793 in type III procollagen in
RT Ehlers-Danlos syndrome type IV.";
RL Hum. Mutat. 5:179-181(1995).
RN [23]
RP VARIANT EDS-IV ASP-1050.
RX MEDLINE=90037070; PubMed=2808425;
RA Kuivaniemi H., Stolle C.A., Pope F.M., Prockop D.J.;
RT "Single base mutation in the type III procollagen gene that converts
RT the codon for glycine 883 to aspartate in a mild variant of
RT Ehlers-Danlos syndrome IV.";
RL J. Biol. Chem. 264:19313-19317(1989).
RN [24]
RP VARIANT EDS-IV VAL-1077.
RX MEDLINE=91374480; PubMed=1895316;
RA Richards A.J., Lloyd J.C., Ward P.N., de Paepe A., Narcisi P.,
RA Pope F.M.;
RT "Characterisation of a glycine to valine substitution at amino acid
RT position 910 of the triple helical region of type III collagen in a
RT patient with Ehlers-Danlos syndrome type IV.";
RL J. Med. Genet. 28:458-463(1991).
RN [25]
RP VARIANT EDS-IV GLU-1173.
RX MEDLINE=93022543; PubMed=1357232;
RA Johnson P.H., Richards A.J., Lloyd J.C., Pope F.M., Hopkinson D.A.,

Query Match
Best Local Similarity 27.6%; Pred. No. 0.055;
Matches 61; Conservative 13; Mismatches 74; Indels 73; Gaps 10;

Score 141.5; DB 1; Length 1466;

```
Qy  171 PRGGAVPRKPVSEMEMERDGA-----AHSLEPGKENLPGDPTSNAT------SRGAEG 218
         |  |   | :    :  ||          :|| || |    : |        ||| |
Db  844 PPGGSGPAGPPGPQGVKGERSSPGGPGCAAGPPGARGLRPGPPGSNDPGPPGPSGSPGKDG 903

Qy  219 PLPP------PSMPAVAGAAGGL-----ALILLGVAGAGGAMCWRRRRA 256
         | ||      || | :|| | :|     | | | ||| || |
Db  904 PPGPAGNTGAPGSPGVSGPKGDAGPDPGEKSSPGAGQPPGAPGPLGIAGITGARGLAGPPG 963

Qy  257 KPSESRHPGP-----GSFGRGGSLGLGG-GGGMGP------REAEPGELGIAL 297
         || |  :|||     ||  || |  | | ||| ||      | |||| |
Db  964 MPGPRGSPGPQGVKGSSGKPGANGLSGERGPGPPGQGLPGLAGTAGEPGRDGNPGSDGLPG 1023

Qy  298 RGG-------GAADPPFCPHYEKVSGDYGHPVYIVQDGPP 330
         |||       || ||    :    |  | |    | |||
Db 1024 RDGSPGGGKGDRGENGSPGAP-----GAPGHP-----GPP 1052
```

ID   CA18_RABIT   STANDARD;   PRT;   744 AA.
AC   P14282;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   Collagen alpha 1(VIII) chain precursor (Endothelial collagen).
GN   COL8A1.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX   NCBI_TaxID=9986;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89380199; PubMed=2476437;
RA   Yamaguchi N., Benya P.D., van der Rest M., Ninomiya Y.;
RT   "The cloning and sequencing of alpha 1(VIII) collagen cDNAs
RT   demonstrate that type VIII collagen is a short chain collagen and
RT   contains triple-helical and carboxyl-terminal non-triple-helical
RT   domains similar to those of type X collagen.";
RL   J. Biol. Chem. 264:16022-16029(1989).
CC   -!- FUNCTION: MAJOR COMPONENT OF THE DESCEMET'S MEMBRANE (BASEMENT
CC   MEMBRANE) OF CORNEAL ENDOTHELIAL CELLS.
CC   -!- SUBUNIT: MAY FORM HOMOTRIMERS, OR HETEROTRIMERS IN ASSOCIATION
CC   WITH ALPHA 2(VIII) TYPE COLLAGENS.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC   UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC   -!- MISCELLANEOUS: FOUR CONSECUTIVE GLY-PRO-PRO TRIPLETS ARE PRESENT
CC   AT THE C-TERMINUS OF THE TRIPLE-HELICAL REGION. THESE MAY PROVIDE
CC   HIGH THERMAL STABILITY OF THIS REGION OF THIS REGION.
CC   -!- SIMILARITY: STRONG, TO ALPHA 2 TYPES VIII AND X COLLAGENS.
CC   -!- SIMILARITY: CONTAINS 1 C1Q DOMAIN.
CC   --
CC   --
DR   EMBL; J05042; AAA31204.1; -.
DR   PIR; A34246; A34246.
DR   InterPro; IPR001073; C1q.
DR   InterPro; IPR000087; Collagen.
DR   Pfam; PF00386; C1q; 1.
DR   Pfam; PF01391; Collagen; 8.
DR   PRINTS; PR00007; COMPLEMNTC1Q.
DR   SMART; SM00110; C1Q; 1.
DR   PROSITE; PS01113; C1Q; 1.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Glycoprotein; Cell adhesion; Collagen; Signal.
FT   SIGNAL   1   20
FT   CHAIN   21   744   COLLAGEN ALPHA 1(VIII) CHAIN.
FT   DOMAIN   29   117   NONHELICAL REGION (NC2).
FT   DOMAIN   118   571   TRIPLE-HELICAL REGION.
FT   DOMAIN   572   744   NONHELICAL REGION (NC1).
FT   DOMAIN   609   744   C1Q.
SQ   SEQUENCE   744 AA; 73358 MW; 2A8CEF1EF82744E99 CRC64;

Query Match
Best Local Similarity 23.6%; Pred. No. 0.029;
Matches 83; Conservative 29; Mismatches 108; Indels 132; Gaps 19;

Score 141; DB 1; Length 744;

```
Qy   67 PPGPHSSSNYEFYKLYLVG----------GAGG-RCEAPAPNLLTCDRPDLDLRFT 114
         ||||| :    | :|  |            | ||            :|  |  | :
Db  198 PPGPHGLPG-----------IGKPGGPGLPGQPGAKGDRGPKGPPGP---------- 233

Qy  115 IKFQEYSPNIWGHEFRSHHDYYIIATSDGTR----EGLESLQG--GVCLTRGMKVLLRVG 168
         |  :|   |:||| ||     :    |        |||:|||    |  | | |  | ||
Db  234 ------PGLQGPK------GEKGFGMPGLPGLKGQPPGMIGPPGPVGLPGVG 272

Qy  169 Q------SPRGGAVPRKPVSEMPMERDGAAHSLEPGKENLPGDP--TSN 210
         |       ||  |                :|| || |    |    |:|
```

RESULT 45

ID   CA18_MOUSE   STANDARD;   PRT;   743 AA.
AC   Q00780; Q9D2V4;
DT   01-NOV-1995 (Rel. 32, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Collagen alpha 1(VIII) chain precursor.
GN   COL8A1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c;
RX   MEDLINE=92362626; PubMed=1499564;
RA   Muragaki Y., Shiota C., Inoue M., Ooshima A., Olsen B.R.,
RA   Ninomiya Y.;
RT   "Alpha 1(VIII)-collagen gene transcripts encode a short-chain
RT   collagen polypeptide and are expressed by various epithelial,
RT   endothelial and mesenchymal cells in newborn mouse tissues.";
RL   Eur. J. Biochem. 207:895-902(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Kidney;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; X66976; CAA47387.1; -.
DR   EMBL; X66977; CAA47387.1; JOINED.
DR   EMBL; AK018742; BAB31383.1; -.
DR   PIR; S23779; S23779.
DR   MGD; MGI:88463; Col8a1.
DR   InterPro; IPR001073; C1q.
DR   InterPro; IPR000087; Collagen.
DR   Pfam; PF00386; C1q; 1.
DR   Pfam; PF01391; Collagen; 7.
DR   PRINTS; PR00007; COMPLEMNTC1Q.
DR   SMART; SM00110; C1Q; 1.
DR   PROSITE; PS01113; C1Q; 1.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Glycoprotein; Cell adhesion; Collagen; Signal.
FT   SIGNAL   1   28   BY SIMILARITY.
FT   CHAIN   29   743   COLLAGEN ALPHA 1(VIII) CHAIN.
FT   DOMAIN   29   117   NONHELICAL REGION (NC2).
FT   DOMAIN   118   571   TRIPLE-HELICAL REGION (COL1).
FT   DOMAIN   572   743   NONHELICAL REGION (NC1).
FT   CONFLICT   6   6   G -> R (IN REF. 2).
FT   CONFLICT   85   85   H -> Y (IN REF. 1).
FT   CONFLICT   109   109   K -> KG (IN REF. 2).
FT   CONFLICT   248   248   P -> L (IN REF. 1).
FT   CONFLICT   313   313   P -> A (IN REF. 1).
FT   CONFLICT   323   324   IP -> SR (IN REF. 1).
FT   CONFLICT   361   361   D -> H (IN REF. 1).
FT   CONFLICT   361   361   D -> H (IN REF. 2).
FT   CONFLICT   596   596   T -> P (IN REF. 2).
FT   CONFLICT   717   719   NPF (IN REF. 1).
SQ   SEQUENCE   743 AA; 73454 MW; F584D85ED53897F4 CRC64;

CC   -!- FUNCTION: MAJOR COMPONENT OF THE DESCEMET'S MEMBRANE (BASEMENT
CC   MEMBRANE) OF CORNEAL ENDOTHELIAL CELLS.
CC   -!- SUBUNIT: MAY FORM HOMOTRIMERS, OR HETEROTRIMERS IN ASSOCIATION
CC   WITH ALPHA 2(VIII) TYPE COLLAGENS.
CC   -!- TISSUE SPECIFICITY: HIGH LEVELS IN CALVARIUM, EYE & SKIN OF
CC   NEWBORN MICE; ALSO IN VARIOUS EPITHELIAL, ENDOTHELIAL AND
CC   MESENCHYMAL CELLS.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC   UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC   -!- SIMILARITY: STRONG, TO ALPHA 2 TYPES VIII AND X COLLAGENS.
CC   -!- SIMILARITY: CONTAINS 1 C1Q DOMAIN.

Query Match   5.7%; Score 140.5; DB 1; Length 743;
Best Local Similarity   21.3%;
Matches 101; Conservative 30; Mismatches 135; Indels 209; Gaps 21;

```
Qy   14 GALLLLGVILGLVSGLSLEPVYWNSANKRFQAEGGYV----LYPQIGDRLDLCPRARPPG 69
         |:||  :|| |:|  : :                         :|  |||| || | ||
Db    6 GPLQILGILFIS-------------LNSV--RLIQAGAYYGIXKPLPPQIPPQIPQYQPLG 55

Qy   70 ---PH--------------SSPNYEYKLY------- 83
           ||              : |  : |
Db   56 QQVPHMPLGKDGLSMGKEMPHMQYGKEYPHLPQYMKEIPPVPRMGKEVVPKKGKEVPLAS 115

Qy   84 VGGAQGRRCB----APPAPNLILTCDRPDLDLRRTIKFQEYSPNIWGHHFRSHHDYYIIA 139
         : | :|  |     |   :|:| :                   |
Db  116 LRGEQGPRGEPGPRGPGP---------------PGLPGHGMPG------IK 146

Qy  140 TSDGT-------REGLESLGG--GVCLIRGMKVLLRVGQSPRGGAVPRKPVSEMPMERDR 190
          :||          |  | ||    | ||            | |   |
Db  147 GKPGEPGGYPGIGKPGMPGMPGKPGAMGMPGAK--------GEIGPKGEIGPNGIPGPQ 196

Qy  191 G--AAHSL----EPGKENLPGDPTSNATSRGAEGPLPPPSM---------- 225
         |    ||     |||:| |||   |  : |||||| |  |
Db  197 GPPGPHGLPGIGIGKPGGPGLPGQPGCAKG-ERGPKGPPGPGLQGPKGEKGFGMPGLPGLKG 255

Qy  226 -PAVAGAAGGLALLLLGVAGAGAGAMCWRRRARKPSESRHPG---------- 265
          |     |||| ||||||                        |||
Db  256 PPGMGGPPGPVGLPGVGKPGVTGPPGPGCGPLGKGPGPPGEPGPQGLIGVPGVQGPPGMPGV 315

Qy  266 ----------------------PGSFGRGGSLGLGG-GGGMGPRE 287
                               |||:|| || ||| ||  |||:|
Db  316 GKPGQDGIPGQPGPPGGKGEBGQGLPGLPGGVPGLPGVGKPGFPGPGKGDRGIGGVPGVLGPR- 374

Qy  288 AEPGELGIALRGGGAADPPFCPHYEKVSGDYGHPVYIVQDGPP----------PQSPP 334
```

```
Db  273 KPGVTGPPGPPGQGPPLGKGPGPPEPGPGGIGVPGVQGPPGPLPGVGPKGPGPGQDGIPGQPGFPGG 332
Qy  211 ATSRGAEGPLPPPSMPAVA------GAAGGLALLLLGVAGAGAGAMCWRRRR----------A 256
Db  333 KCGEQGLPGPPGLPGVGKPGFPGPGKGD----RGIGGVPGALGPPGRGKPGVGAPGMGG 387
Qy  257 KPSESRHPG-PGSFGRGGSLGL----GGGGGGGMGPR------EAEPGELGIALRGG--GAAD 304
Db  388 PPGEPGLPGIPGPMGPPGPGAIGPPGPKGEGGIVGPQGPPGKGEPGIPGPGPGPKGPGFLGEVG 447
Qy  305 PP---FCPHYEKVSGDYGH-------PVVIVQDGPP------QSPPNI 336
Db  448 PPGIRGLPGPGPPGPKGEAGHKGLPGLPGGVPGLLGPKGPGEPGIPGDQGILQGPPGI 499
```

```
Db

  RESULT 46

  375 GEKGPIGPAPGMGG-----PPGEPGLPGIPGPMGPPGAIGPPGPKGEGGVVGPQGPP 425
       ||| ||    |||          : |||: : ||| ||:||  |  | |  ||:  |||

ID   CA1A_HUMAN    STANDARD;      PRT;   680 AA.
AC   Q03692;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Collagen alpha 1(X) chain precursor.
GN   COL10A1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92109659; PubMed=1764025;
RA   Thomas J.T., Cresswell C.J., Rash B., Nicolai H., Jones T.,
RA   Solomon E., Grant M.E., Boot-Handford R.P.;
RT   "The human collagen X gene. Complete primary translated sequence and
RT   chromosomal localization.";
RL   Biochem. J. 280:617-623(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93012005; PubMed=1397333;
RA   Reichenberger E., Beier F., Luvalle P., Olsen B.R., von der Mark K.,
RA   Bertling W.M.;
RT   "Genomic organization and full-length cDNA sequence of human collagen
RT   X.";
RL   FEBS Lett. 311:305-310(1992).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Beier F., Lammi M.B., von der Mark K.;
RL   Submitted (JUN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Williams S.;
RL   Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE OF 52-680 FROM N.A.
RX   MEDLINE=92267014; PubMed=1587271;
RA   Apte S.S., Seldin M.F., Hayashi M., Olsen B.R.;
RT   "Cloning of the human and mouse type X collagen genes and mapping of
RT   the mouse type X collagen gene to chromosome 10.";
RL   Eur. J. Biochem. 206:217-224(1992).
RN   [6]
RP   SEQUENCE OF 547-655 FROM N.A.
RX   MEDLINE=91243838; PubMed=2037056;
RA   Apte S., Mattei M.-G., Olsen B.R.;
RT   "Cloning of human alpha 1(X) collagen DNA and localization of the
RT   COL10A1 gene to the q21-q22 region of human chromosome 6.";
RL   FEBS Lett. 282:393-396(1991).
RN   [7]
RP   SEQUENCE OF 561-666 FROM N.A.
RX   MEDLINE=91012290;
RA   Reichenberger E., Aigner T., von der Mark K., Stoeb H., Bertling W.;
RT   "In situ hybridization studies on the expression of type X collagen
RT   in fetal human cartilage.";
RL   Dev. Biol. 148:562-572(1991).
RN   [8]
RP   REVIEW ON VARIANTS.
RX   MEDLINE=97255959; PubMed=9101290;
RA   Kuivaniemi H., Tromp G., Prockop D.J.;
RT   "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
RT   associated collagen (type IX), and network-forming collagen (type X)
RT   cause a spectrum of diseases of bone, cartilage, and blood vessels.";
RL   Hum. Mutat. 9:300-315(1997).
RN   [9]
RP   VARIANTS SMCD ASP-598 AND PRO-614.
RX   MEDLINE=94136476; PubMed=8304336;
```

```
RA   Wallis G.A., Rash B., Sweetman W.A., Thomas J.T., Super M., Evans G.,
RA   Grant M.E., Boot-Handford R.P.;
RT   "Amino acid substitutions of conserved residues in the
RT   carboxyl-terminal domain of the alpha 1(X) chain of type X collagen
RT   occur in two unrelated families with metaphyseal chondrodysplasia
RT   type Schmid.";
RL   Am. J. Hum. Genet. 54:169-178(1994).
RN   [10]
RP   VARIANT SMCD ARG-591.
RX   MEDLINE=94272470; PubMed=8004099;
RA   McIntosh I., Abbott M.H., Warman M.L., Olsen B.R., Francomano C.A.;
RT   "Additional mutations of type X collagen confirm COL10A1 as the
RT   Schmid metaphyseal chondrodysplasia locus.";
RL   Hum. Mol. Genet. 3:303-307(1994).
RN   [11]
RP   VARIANT SMCD VAL-618.
RX   MEDLINE=95181449; PubMed=7876225;
RA   Chan D., Cole W.G., Rogers J.G., Bateman J.F.;
RT   "Type X collagen multimer assembly in vitro is prevented by a Gly618
RT   to Val mutation in the alpha 1(X) chain of type X collagen resulting in Schmid
RT   metaphyseal chondrodysplasia.";
RL   J. Biol. Chem. 270:4558-4562(1995).
RN   [12]
RP   VARIANTS SMCD ARG-545; GLU-595; HIS-597; LYS-617; ARG-644 AND GLY-648.
RX   MEDLINE=95331767; PubMed=7607655;
RA   Bonaventure J., Chaminade F., Maroteaux P.;
RT   "Mutations in three subdomains of the carboxy-terminal region of
RT   collagen type X account for most of the Schmid metaphyseal
RT   dysplasias.";
RL   Hum. Genet. 96:58-64(1995).
RN   [13]
RP   VARIANT SMCD PRO-600.
RX   MEDLINE=96375754; PubMed=8782043;
RA   Wallis G.A., Rash B., Sykes B., Bonaventure J., Maroteaux P.,
RA   Zabel B., Wynne-Davies R., Grant M.E., Boot-Handford R.P.;
RT   "Mutations within the gene encoding the alpha 1 (X) chain of type X
RT   collagen (COL10A1) cause metaphyseal chondrodysplasia type Schmid but
RT   not several other forms of metaphyseal chondrodysplasia.";
RL   J. Med. Genet. 33:450-457(1996).
RN   [14]
RP   VARIANTS SMCD GLU-18 AND ARG-18.
RX   MEDLINE=97220591; PubMed=9067753;
RA   Ikegawa S., Nakamura K., Nagano A., Haga N., Nakamura Y.;
RT   "Mutations of the type X collagen gene (COL10A1) causes
RT   spondylometaphyseal dysplasia.";
RL   Am. J. Hum. Genet. 63:1659-1662(1998).
RN   [15]
RP   VARIANT SMD GLU-595.
RX   MEDLINE=99057503; PubMed=9837818;
RA   Ikegawa S., Nishimura G., Nagai T., Hasegawa T., Ohashi H.,
RA   Nakamura Y.;
RT   "Mutation of the type X collagen gene 'COL10A1' causes
RT   spondylometaphyseal dysplasia.";
RL   Hum. Mutat. 9:131-135(1997).
RN   [16]
RP   VARIANT SMCD CYS-597.
RX   MEDLINE=99069781; PubMed=9852679;
RA   Sawai H., Ida A., Nakata Y., Koyama K.;
RT   "Novel missense mutation resulting in the substitution of tyrosine by
RT   cysteine at codon 597 of the type X collagen gene associated with
RT   Schmid metaphyseal chondrodysplasia.";
RL   J. Hum. Genet. 43:259-261(1998).
CC   -!- FUNCTION: TYPE X COLLAGEN IS A PRODUCT OF HYPERTROPHIC
CC       CHONDROCYTES AND HAS BEEN LOCALIZED TO PRESUMPTIVE
CC       MINERALIZATION ZONES OF HYALINE CARTILAGE.
CC   -!- SUBUNIT: HOMOTRIMER.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC       UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC   -!- DISEASE: DEFECTS IN COL10A1 ARE THE CAUSE OF SCHMID TYPE
CC       METAPHYSEAL CHONDRODYSPLASIA (SMCD), A DOMINANTLY INHERITED
CC       DISORDER OF THE OSSEOUS SKELETON. THE CARDINAL FEATURES OF THE
CC       PHENOTYPE ARE MILD SHORT STATURE, COXA VARA AND A WADDLING GAIT.
```

CC   RADIOGRAPHY USUALLY SHOWS SCLEROSIS OF THE RIBS, FLARING OF THE
CC   METAPHYSES, AND A WIDE IRREGULAR GROWTH PLATE, ESPECIALLY OF THE
CC   KNEES.
CC   -!- DISEASE: DEFECTS IN COL10A1 ARE THE CAUSE OF SPONDYLOMETAPHYSEAL
CC   DYSPLASIA (SMD). SMD COMPRISES A HETEROGENEOUS GROUP OF HERITABLE
CC   SKELETAL DYSPLASIAS CHARACTERIZED BY MODIFICATIONS OF THE
CC   VERTEBRAL BODIES OF THE SPINE AND METAPHYSES OF THE TUBULAR BONES.
CC   -!- SIMILARITY: STRONG, TO ALPHA 1 AND 2 TYPE VIII COLLAGENS.
CC   -!- SIMILARITY: CONTAINS 1 C1Q DOMAIN.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; X60382; CAA42933.1; -.
DR   EMBL; X65120; CAA46236.1; -.
DR   EMBL; X98658; CAA67178.1; -.
DR   EMBL; AL121963; CAB87590.1; -.
DR   EMBL; S68531; AAC60615.1; -.
DR   EMBL; X58879; CAA41686.1; -.
DR   EMBL; M74050; AAA61221.1; -.
DR   EMBL; X72580; CAA51170.1; JOINED.
DR   EMBL; S15826; S15826.
DR   PIR; S30086; S30086.
DR   PIR; A43901; A43901.
DR   PIR; S18849; S18849.
DR   PIR; S21856; S21856.
DR   PIR; S26396; S26396.
DR   Genew; HGNC:2185; COL10A1.
DR   MIM; 120110; -.
DR   MIM; 156500; -.
DR   MIM; 184250; -.
DR   InterPro; IPR001073; C1q.
DR   InterPro; IPR000087; Collagen.
DR   Pfam; PF00386; C1q; 1.
DR   Pfam; PF01391; Collagen; 8.
DR   PRINTS; PR00007; COMPLEMNTC1Q.
DR   SMART; SM00110; C1Q; 1.
DR   PROSITE; PS01113; C1Q; 1.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Cartilage; Collagen; Signal; Disease mutation; Polymorphism.
FT   SIGNAL   1   18   POTENTIAL.
FT   CHAIN   19   680   COLLAGEN ALPHA 1(X) CHAIN.
FT   DOMAIN   19   56   NONHELICAL REGION (NC2).
FT   DOMAIN   57   519   TRIPLE-HELICAL REGION.
FT   DOMAIN   520   680   NONHELICAL REGION (NC1).
FT   DOMAIN   545   680   C1Q.
FT   VARIANT   18   18   G -> E (IN SMCD).
FT   /FTId=VAR_001838.
FT   VARIANT   18   18   G -> R (IN SMCD).
FT   /FTId=VAR_001839.
FT   VARIANT   545   545   G -> R.
FT   /FTId=VAR_001840.
FT   VARIANT   591   591   C -> R (IN SMCD).
FT   /FTId=VAR_001841.

Query Match   5.7%; Score 140; DB 1; Length 680;
Best Local Similarity 29.1%; Pred. No. 0.031;
Matches   57; Conservative   13; Mismatches   84; Indels   42; Gaps   9;

Qy   171 PRG-----GAVPRKP----------VSEMPMERDRGAAHSLEPGKENLPG-------DPTSNA 211

Db   70 PRGHPGPSGGPPGKPGYGSPGCGLQGEPGLPGPGPGPSAVGKPGVPGLPGKPGERGPYGPKGDV 129

Qy   212 TSRGAEGPLPPPSMPAVAGAAGGLALLLGVAGAGGAMCWRRRARAKPSESRHPG-PGSFG 270

Db   130 GPAGLPGPRGPPGPGIPGPAG---ISVPGKPGGQGPTGAPGPRGPRGFPGEKGAPGVPGMNG 186

Qy   271 RGGSLGLGGGGMGPREAFPGELGITALRGG--GAADPP------FCPHYEKVSGDYGH 320

Db   187 QKGEMGYGAPG------RPGERGLPGPGQGPTGPSGPPGVGKRGENGVPGQPGIKGDRGF 239

Qy   321 PVYIVQDGP--PQSPP 334

Db   240 PGEMGPIGPPGPQGPP 255

RESULT 47
CAC2_HAECO

ID   CAC2_HAECO   STANDARD;   PRT;   210 AA.
AC   P16252;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-AUG-1990 (Rel. 15, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Cuticle collagen 2C (Fragment).
GN   2C.
OS   Haemonchus contortus (Barber pole worm).
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;
OC   Trichostrongyloidea; Haemonchidae; Haemonchinae; Haemonchus.
OX   NCBI_TaxID=6289;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90136718; PubMed=2615789;
RA   Shamansky L.M., Pratt D., Boisvenue R.J., Cox G.N.;
RT   "Cuticle collagen genes of Haemonchus contortus and Caenorhabditis
RT   elegans are highly conserved.";
RL   Mol. Biochem. Parasitol. 37:73-86(1989).
CC   -!- FUNCTION: NEMATODE CUTICLES ARE COMPOSED LARGELY OF COLLAGEN-LIKE
CC   PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A
CC   BARRIER TO PROTECT THE WORM FROM ITS ENVIRONMENT.
CC   -!- MISCELLANEOUS: THIS PROTEIN SHOWS 4 POTENTIAL TRIPLE-HELICAL
CC   REGIONS, WHICH CONTAIN GLYCINE AS EVERY THIRD AMINO ACID.
CC   -!- MISCELLANEOUS: IN ALL NEMATODE CUTICLE COLLAGENS, THE POLYPEPTIDE
CC   CHAINS ARE COMPLEXED WITHIN THE CUTICLE BY DISULFIDE BONDS AND
CC   OTHER TYPES OF COVALENT CROSS-LINKS.
CC   -!- SIMILARITY: BELONGS TO THE CUTICULAR COLLAGEN FAMILY.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; J04670; AAA29172.1; -.
DR   InterPro; IPR000087; Collagen.
DR   Pfam; PF01391; Collagen; 3.
KW   Cuticle; Connective tissue; Repeat; Multigene family; Collagen.
FT   NON_TER   1   1
SQ   SEQUENCE   210 AA;   19562 MW; E15FAA9A2DF3D74B CRC64;

Query Match   5.7%; Score 139.5; DB 1; Length 210;
Best Local Similarity 31.1%; Pred. No. 0.01;
Matches   50; Conservative   5; Mismatches   75; Indels   31; Gaps   5;

Qy   176 VPRKPVSEMPMERDRGAAHSLEPGKENLPGDP--TSNATSRGAEGPLPPPSMPAVAGAAG 233

Db   51 IPIPPCKPCPQGPRGPPGPPGIPGPAGPGTGNPGAPGPKPGKPGPPGPGKPGAG- 109

Qy   234 GLALLLGVAGAGGAMCWRRRARAKPSESRHPG-----PGSFGRGGSLGLGGGGGMGPREAE 289

Db   110 -------GAAGQPGANA-----PSEPLVPGPGPGPPGPTGPEGPPGPNGAPGHPGAPGA 155

Qy   290 PGELGITALRGGGAADPPFCPHYEKVSGDYGHPVYIVQDGPP 330

Db   156 PGEKGPRGQDGHPGAP--------GNAGIPGGPGQPGPP 186

RESULT 48
CA1B_MOUSE

```
ID   CA1B_MOUSE     STANDARD;      PRT;   1804 AA.
AC   Q61245; Q64047;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Collagen alpha 1(XI) chain precursor.
GN   COL11A1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.
RC   STRAIN=C57BL/6;
RC   TISSUE=Embryo;
RX   MEDLINE=95163095; PubMed=7859283;
RA   Li Y., Lacerda D.A., Warman M.L., Beier D.R., Yoshioka H.,
RA   Ninomiya Y., Oxford J.T., Morris N.P., Andrikopoulos K.,
RA   Ramirez F., Wardell B.J., Lifferth G.D., Teuscher C., Woodward S.R.,
RA   Taylor B.A., Seegmiller R.E., Olsen B.R.;
RT   "A fibrillar collagen gene, Col11a1, is essential for skeletal
RT   morphogenesis.";
RL   Cell 80:423-430(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96015067; PubMed=8530046;
RA   Yoshioka H., Inoguchi K., Khaleduzzaman M., Ninomiya Y.,
RA   Andrikopoulos K., Ramirez F.;
RT   "Coding sequence and alternative splicing of the mouse alpha 1(XI)
RT   collagen gene (Col11a1).";
RL   Genomics 28:337-340(1995).
CC   -!- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN FIBRILLOGENESIS BY
CC       CONTROLLING LATERAL GROWTH OF COLLAGEN II FIBRILS.
CC   -!- SUBUNIT: TRIMERS COMPOSED OF THREE DIFFERENT CHAINS: ALPHA 1(XI),
CC       ALPHA 2(XI), AND ALPHA 3(XI). ALPHA 3(XI) IS A POST-TRANSLATIONAL
CC       MODIFICATION OF ALPHA 1(II) (BY SIMILARITY).
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC       SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC       UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC   -!- DISEASE: DEFECTS IN COL11A1 ARE ASSOCIATED WITH CHONDRODYSPLASIA,
CC       AN AUTOSOMAL RECESSIVE DISEASE CHARACTERIZED BY SKELETAL DEFECTS
CC       CAUSED BY ABNORMALITIES IN THE CARTILAGE OF LIMBS, RIBS, MANDIBLES
CC       AND TRACHEA.
CC   -!- SIMILARITY: BELONGS TO THE FIBRILLAR CLASS OF COLLAGENS.
CC   -!- SIMILARITY: HIGH, TO ALPHA 1(V) AND ALPHA 3(V) CHAINS.
CC   --------------------------------------------------------------------------
```

```
DR   EMBL; D38162; BAA07367.1; -.
DR   EMBL; S74574; AAB33439.1; -.
DR   MGD; MGI:88446; Col11a1.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000085; Fib_collagen_C.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR001230; Prenyl_site.
DR   InterPro; IPR003129; TSPN.
DR   Pfam; PF01391; Collagen; 16.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF02210; TSPN; 1.
DR   ProDom; PD000007; Collagen; 1.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00282; LamG; 1.
DR   SMART; SM00210; TSPN; 1.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Glycoprotein; Collagen; Signal; Alternative splicing;
KW   Disease mutation.
FT   SIGNAL        1     35       POTENTIAL.
FT   PROPEP       36            AMINO-TERMINAL PROPEPTIDE (POTENTIAL).
FT   CHAIN       512   1561      COLLAGEN ALPHA 1(XI) CHAIN.
FT   PROPEP     1562   1804      CARBOXYL-TERMINAL PROPEPTIDE.
FT   DOMAIN       37    417      NONHELICAL REGION.
FT   DOMAIN      418    506      TRIPLE-HELICAL REGION (INTERRUPTED).
FT   DOMAIN      507    509      NONHELICAL REGION.
FT   DOMAIN      510    527      TELOPEPTIDE.
FT   DOMAIN      528   1540      TRIPLE-HELICAL REGION.
FT   DOMAIN     1541   1561      TELOPEPTIDE.
FT   DOMAIN     1561   1638      NONHELICAL REGION (C-TERMINAL).
FT   CARBOHYD   1638   1638      N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    610    610      N-LINKED (GLCNAC...) (POTENTIAL).
FT   SITE       1450   1450      CROSSLINKING.
FT   SITE        610    610      CROSSLINKING.
FT   VARSPLIC    329    413      MISSING (IN SHORT ISOFORM).
FT   VARIANT     189    195      LDRSERS -> SIEVRDR (IN CHONDRODYSPLASIA).
FT   VARIANT     196    196      MISSING (IN CHONDRODYSPLASIA).
SQ   SEQUENCE   1804 AA;  180963 MW;  FE2DB9DEDIE42I9A CRC64;

Query Match   5.7%;  Score 139;  DB 1;  Length 1804;
Best Local Similarity   25.7%;  Pred. No. 0.099;
Matches  105;  Conservative  17;  Mismatches  116;  Indels 170;  Gaps 24;

Qy     2 GPP-HSGPGGGVRVGALLLGVLGL--VSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRL 58
           |||  |    |          |    |||
Db   440 GPPGPAGPAG------LMGPPGILQGPSGLPGDD----------------------GDR- 469

Qy    59 DLLCPRARPPGPHSSPNYEFYKLYLVGPAQGRRCEAPPAPNLLLTCDRPDLDLRF------ 113
           |||
Db   470 -------GPPGRPGLP----------GADG--LPGPPGTMLM--------LPFRYGGDGS 502

Qy   114 ---TIKFQEYSPNLMGHEFRSHHDYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVG-- 168
           ||
Db   503 KGPTISAQERAQAQ-----------------AILQQARIALRGPPGPMGLTG------RPGPV 541

Qy   169 ------QSPRGGAVPRKPVSE-----------MPMER----DRG-- 191
Db   542 GGPGSTGAKGESGDPGPGPGPGRGVQGPPGPPGPGKPGKRGPGADGPGRGMPGPGSGSKGDRGFD 601

Qy   192 ----AAHSLPGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAGGLALLLLGVA 243
Db   602 GLPGLPGDKGHRGERGPGPGPPGPLGDGDG--MRGEDGEIGPRGLPGEAGPRG----LIGPR 655

Qy   244 GA--------GGAVCWRRRRAKPSESRHFG-PGSFGRGGSLGL-GGGGGMGPRRAEPGE 292
Db   656 GTPGPPGQPGIGGIDGPGFKGNNMGPQGEPGPPGQGNPGSQGLPGPQGPIGP----PGE 711

Qy   293 LGIALRGG-----GAADPPPCPHYEKVSGDYGHPVYIVQDGP--PQSP 333
Db   712 KGPQGRPGLAGLPGDADGPGPGHPGKEGGGSGEKG-----ALGPPGPQGP 753

RESULT 49
ID   CA12_MOUSE     STANDARD;      PRT;   1459 AA.
AC   P28481;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Collagen alpha 1(II) chain precursor [Contains: Chondrocalcin].
GN   COL2A1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91358489; PubMed=1885613;
RA   Metsaranta M., Toman D., de Crombrugghe B., Vuorio E.;
RT   "Mouse type II collagen gene. Complete nucleotide sequence, exon
RT   structure, and alternative splicing.";
RL   J. Biol. Chem. 266:16862-16869(1991).
```

```
RN   [2]
RP   SEQUENCE OF 1455-1459 FROM N.A.
RX   MEDLINE=91274355; PubMed=2054384;
RA   Metsaranta M., Toman D., de Crombrugghe B., Vuorio E.;
RT   "Specific hybridization probes for mouse type I, II, III and IX
RT   collagen mRNAs.";
RL   Biochim. Biophys. Acta 1089:241-243(1991).
CC   -!- FUNCTION: COLLAGEN TYPE II IS SPECIFIC FOR CARTILAGINOUS TISSUES.
CC   -!- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(II) CHAINS.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC       UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC   -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC   -----------------------------------------------------------------
```

```
DR   EMBL; M65161; AAA68100.1; -.
DR   EMBL; X57982; CAA41047.1; -.
DR   MGD; MGI:88452; Col2a1.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001007; VWF_C.
DR   Pfam; PF00093; vwc; 1.
DR   Pfam; PF01391; Collagen; 18.
DR   Pfam; PF01410; COLFI; 1.
DR   ProDom; PD000007; Collagen; 3.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS01208; VWFC; 1.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Glycoprotein; Collagen; Cartilage; Signal; Alternative splicing.
FT   SIGNAL        1     25       POTENTIAL.
FT   PROPEP       26    153       AMINO-TERMINAL PROPEPTIDE
FT   CHAIN       154   1213       COLLAGEN ALPHA 1(II) CHAIN.
FT   PROPEP     1214   1459       CARBOXYL-TERMINAL PROPEPTIDE
FT                                (CHONDROCALCIN).
FT   DOMAIN       32     89       VWFC.
FT   DOMAIN      173   1186       TRIPLE-HELICAL REGION.
FT   DOMAIN     1187   1213       NONHELICAL REGION (C-TERMINAL).
FT   VARSPLIC     29     29       Q -> R (IN SHORT ISOFORM).
FT   VARSPLIC     30     98       MISSING (IN SHORT ISOFORM).
SQ   SEQUENCE   1459 AA;  139154 MW;  F6C84FA7C5S32E7E2 CRC64;

Query Match             5.7%;  Score 138.5;  DB 1;  Length 1459;
Best Local Similarity   30.1%;  Pred. No. 0.086;
Matches   58;  Conservative  13;  Mismatches  75;  Indels  47;  Gaps  9;

Qy   173 GGAVPRKVSEMPMERDGAAAHSLEPGKENL------PGD--PTSNAITSRGABGPLPPPS 224
             ||:|                            ||:|   |:|||:||||:|||||
Db   470 GGAGPIGPPGERGAPGNRGF-----PGQDGLAGPKGAPGERGPSGLAGPKGANGDPGRPG 524

Qy   225 MPAVAGAAGGLIALILLIGVAGAGAGAMCWRRRAKPSESRHPGP------GSFGRGGSIGLGGG 280
Db   525 EPGLPGARG-----LTGRPGDAGPQGKVGPSGAPGEDGRPGPGPQGARGQGVMGFPGP 579

Qy   281 GGMGPREAEPGELGIA-----LRG-------GGAADPPFCPHYEKVSGDYGHPVYIVQDGP 329
Db   580 KGANGEPGPKAGEKGLAGAPGLRGLPGKDGETGAAGPPGPSGPSGPAGERGEQGAP------GP 633

Qy   330 ------PQSPP 334
             ||||
Db   634 SGFPGGLPGPGPGP 646
```

```
ID   CA2B_HUMAN     STANDARD;     PRT;  1736 AA.
AC   P13942; Q13273; Q13271; Q13272; Q07751; Q99866; Q9UIP9;
DT   01-JAN-1990 (Rel. 13, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Collagen alpha 2(XI) chain precursor.
GN   COL11A2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96032717; PubMed=7559422;
RA   Vuoristo M.M., Pihlajamaa T., Vandenberg P., Prockop D.J.,
RA   Ala-Kokko L.;
RT   "The human COL11A2 gene structure indicates that the gene has not
RT   evolved with the genes for the major fibrillar collagens.";
RL   J. Biol. Chem. 270:22873-22881(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Tubby B.;
RL   Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 59-807 FROM N.A.
RC   TISSUE=Cartilage;
RX   MEDLINE=93314796; PubMed=8325374;
RA   Zhidkova N.I., Brewton R.G., Mayne R.;
RT   "Molecular cloning of PARP (proline/arginine-rich protein) from human
RT   cartilage and subsequent demonstration that PARP is a fragment of the
RT   NH2-terminal domain of the collagen alpha 2(XI) chain.";
RL   FEBS Lett. 326:25-28(1993).
RN   [4]
RP   SEQUENCE OF 730-1690 FROM N.A.
RX   MEDLINE=89340485; PubMed=2760050;
RA   Kimura T., Cheah K.S.E., Chan S.D.H., Lui V.C.H., Mattei M.-G.,
RA   van der Rest M., Ono K., Solomon E., Ninomiya Y., Olsen B.R.;
RT   "The human alpha 2(XI) collagen (COL11A2) chain. Molecular cloning of
RT   cDNA and genomic DNA reveals characteristics of a fibrillar collagen
RT   with differences in genomic organization.";
RL   J. Biol. Chem. 264:13910-13916(1989).
RN   [5]
RP   SEQUENCE OF 1-537 FROM N.A.
RX   MEDLINE=96435918; PubMed=8838804;
RA   Lui V.C., Ng L.J., Sat E.W., Cheah K.S.;
RT   "The human alpha 2(XI) collagen gene (COL11A2): completion of coding
RT   information, identification of the promoter sequence, and precise
RT   localization within the major histocompatibility complex reveal
RT   overlap with the KE5 gene.";
RL   Genomics 32:401-412(1996).
RN   [6]
RP   ALTERNATIVE SPLICING.
RX   MEDLINE=95238468; PubMed=7721876;
RA   Zhidkova N.I., Justice S.K., Mayne R.;
RT   "Alternative mRNA processing occurs in the variable region of the
RT   pro-alpha 1(XI) and pro-alpha 2(XI) collagen chains.";
RL   J. Biol. Chem. 270:9486-9493(1995).
RN   [7]
RP   DISEASE.
RX   PubMed=10677296;
RA   Melkoniemi M., Brunner H.G., Manouvrier S., Hennekam R.,
RA   Superti-Furga A., Kaeaeriaeinen H., Pauli R.M., van Essen T.,
RA   Warman M.L., Bonaventure J., Miny P., Ala-Kokko L.;
RT   "Autosomal recessive disorder otospondylomegaepiphyseal dysplasia is
RT   associated with loss-of-function mutations in the COL11A2 gene.";
RL   Am. J. Hum. Genet. 66:368-377(2000).
RN   [8]
RP   REVIEW ON VARIANTS.
RX   MEDLINE=97255959; PubMed=9101290;
RA   Kuivaniemi H., Tromp G., Prockop D.J.;
RT   "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
RT   associated collagen (type IX), and network-forming collagen (type X)
RT   cause a spectrum of diseases of bone, cartilage, and blood vessels.";
```

```
RESULT 50
CA2B_HUMAN
```

```
RL   Hum. Mutat. 9:300-315(1997).
RN   [9]
RP   VARIANT OSMED ARG-661.
RX   MEDLINE=95163096; PubMed=7859284;
RA   Vikkula M., Mariman E.C.M., Lui V.C.H., Zhidkova N.I., Tiller G.E.,
RA   Goldring M.B., van Beersum S.E.C., de Waal Malefijt M.C.,
RA   van den Hoogen F.H.J., Ropers H.-H., Mayne R., Cheah K.S.E.,
RA   Olsen B.R., Warman M.L., Brunner H.G.;
RT   "Autosomal dominant and recessive osteochondrodysplasias associated
RT   with the COL11A2 locus.";
RL   Cell 80:431-437(1995).
RN   [10]
RP   VARIANTS GLY-593; LYS-824; LEU-879; THR-1316 AND GLN-1600.
RX   PubMed=9585596;
RA   Koga H., Sakou T., Taketomi E., Hayashi K., Numasawa T., Harata S.,
RA   Yone K., Matsunaga S., Otterud B., Inoue I., Leppert M.;
RT   "Genetic mapping of ossification of the posterior longitudinal
RT   ligament of the spine.";
RL   Am. J. Hum. Genet. 62:1460-1467(1998).
RN   [11]
RP   VARIANT W2S GLU-955.
RX   PubMed=9505126;
RA   Pihlajamaa T., Prockop D.J., Faber J., Winterpacht A., Zabel B.,
RA   Giedion A., Wiesbauer P., Spranger J., Ala-Kokko L.;
RT   "Heterozygous glycine substitution in the COL11A2 gene in the original
RT   patient with the Weissenbacher-Zweymueller syndrome demonstrates its
RT   identity with heterozygous OSMED (nonocular Stickler syndrome).";
RL   Am. J. Med. Genet. 80:115-120(1998).
RN   [12]
RP   VARIANT STL3 940-GLY-PRO-948 DEL.
RX   PubMed=9506662;
RA   Sirko-Osadsa D.A., Murray M.A., Scott J.A., Lavery M.A., Warman M.L.,
RA   Robin N.H.;
RT   "Stickler syndrome without eye involvement is caused by mutations in
RT   COL11A2, the gene encoding the alpha-2(XI) chain of type XI
RT   collagen.";
RL   J. Pediatr. 132:368-371(1998).
RN   [13]
RP   VARIANTS DFNA13 GLU-808 AND CYS-1034, AND REVISIONS TO 1031-1032.
RX   MEDLINE=20047768; PubMed=10581026;
RA   McGuirt W.T., Prasad S.D., Griffith A.J., Kunst H.P.M., Green G.E.,
RA   Shpargel K.B., Runge C., Huybrechts C., Mueller R.F., Lynch E.,
RA   Brunner H.-C., Brunner H.G., Cremers C.W.R.J., Takanosu M., Li S.-W.,
RA   Arita M., Mayne R., Prockop D.J., Van Camp G., Smith R.J.H.;
RT   "Mutations in COL11A2 cause non-syndromic hearing loss (DFNA13).";
RL   Nat. Genet. 23:413-419(1999).
CC   -!- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN FIBRILLOGENESIS BY
CC       CONTROLLING LATERAL GROWTH OF COLLAGEN II FIBRILS.
CC   -!- SUBUNIT: TRIMERS COMPOSED OF THREE DIFFERENT CHAINS: ALPHA 1(XI),
CC       ALPHA 2(XI), AND ALPHA 3(XI). ALPHA 3(XI) IS A POST-TRANSLATIONAL
CC       MODIFICATION OF ALPHA 1(II). ALPHA 1(V) CAN ALSO BE FOUND INSTEAD
CC       OF ALPHA 3(XI)=1(II).
CC   -!- ALTERNATIVE PRODUCTS: 8 isoforms; 1 (shown here), 2, 3, 4, 5, 6, 7
CC       and 8; may be produced by alternative splicing. They lack exons 6,
CC       7 or 8 or a combination of these exons.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC       UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC   -!- PTM: A DISULFIDE-BONDED PEPTIDE CALLED PROLINE/ARGININE-RICH
CC       PROTEIN OR PARP IS RELEASED FROM THE AMINO TERMINUS DURING
CC       EXTRACELLULAR PROCESSING AND IS SUBSEQUENTLY RETAINED IN THE
CC       CARTILAGE MATRIX FROM WHICH IT CAN BE ISOLATED IN SIGNIFICANT
CC       AMOUNTS.
CC   -!- DISEASE: Defects in COL11A2 are the cause of Stickler syndrome
CC       type 3 (STL3). It is an autosomal dominant disorder characterized
CC       by oro-facial, auditory and skeletal manifestations, such as
CC       midfacial hypoplasia, cleft palate, osteoarthritis, and
CC       sensorineural hearing loss. Differently from Stickler syndrome
CC       type 1 and 2, no ocular involvement is observed. This disorder is
CC       also referred to as Stickler-like syndrome or non-ocular Stickler
CC       syndrome.
CC   -!- DISEASE: Defects in COL11A2 are the cause of autosomal recessive
CC       otospondylomegaepiphyseal dysplasia (OSMED), a skeletal dysplasia
CC       accompanied by severe hearing loss. The phenotype overlaps that of
CC       autosomal dominant skeletal disorders (Stickler and Marshall
CC       syndromes) but can be distinguished by disproportionately short
CC       limbs and lack of ocular involvement.
CC   -!- DISEASE: Defects in COL11A2 are the cause of Weissenbacher-
CC       Zweymueller syndrome (WZS), an autosomal dominant disorder allelic
CC       with STL3 and OSMED. It is also referred to as heterozygous OSMED.
CC   -!- DISEASE: Defects in COL11A2 are the cause of autosomal dominant
CC       nonsyndromic sensorineural deafness type 13 (DFNA13). Affected
CC       individuals experience progressive hearing loss beginning in the
CC       second to fourth decades, eventually making use of amplification
CC       mandatory.
CC   -!- SIMILARITY: BELONGS TO THE FIBRILLAR CLASS OF COLLAGENS.
CC   -!- DATABASE: NAME=Hereditary hearing loss homepage;
CC       NOTE=Gene page;
CC       WWW="http://www.uia.ac.be/dnalab/hhh/hhhgenes.html".
```

```
DR   EMBL; U32169; AAC50214.1; -.
DR   EMBL; U32169; AAC50213.1; -.
DR   EMBL; U32169; AAC50015.1; -.
DR   EMBL; AL031228; CAA20240.1; -.
DR   EMBL; L19987; AAA35498.1; -.
DR   EMBL; J04974; AAA52034.1; -.
DR   EMBL; U41065; AAC17464.1; JOINED.
DR   EMBL; U41065; AAC17464.1; JOINED.
DR   EMBL; U41066; AAC17464.1; JOINED.
DR   EMBL; U41067; AAC17464.1; JOINED.
DR   PIR; A32645; A32645.
DR   Genew; HGNC;2187; COL11A2.
DR   MIM; 120290; -.
DR   MIM; 184840; -.
DR   MIM; 215150; -.
DR   MIM; 215150; -.
DR   MIM; 601868; -.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR001230; Prenyl_site.
DR   InterPro; IPR003129; TSPN.
DR   Pfam; PF01391; Collagen; 18.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF02210; TSPN; 1.
DR   ProDom; PD000007; Collagen; 1.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00282; LamG; 1.
DR   SMART; SM00210; TSPN; 1.
```

```
Query Match    5.7%;  Score 138.5;  DB 1;  Length 1736;
Best Local Similarity   31.7%;
Matches   64;  Conservative   18;  Mismatches   71;  Indels   49;  Gaps   14;

Qy   147  GLESLQGGYVCLLTRGMKVLLRVGQSPRGGAVPRK--PVSEMPMERDRGAAHSLRPGKENLP 204
               |||   |         |   |  |  |     |         |    |   |    |  |
Db  1327  GSEGRQGG-------KGAK-----------GDPGAIGAPGKTGPVGP-----AGPAGKPGPDGLR 1368

Qy   205  GDPTS------NATSR-GAEGPLPPPSMPAVAGAAGGLALL----LLGVAGAGGAMCWR 252
               || |        |  |  || ||        ||| |  ||||      || | || | |
Db  1369  GLPGSVGQQGQRPGAIGQAGPPGPVGPPGLPGLRGDAGAKGEKGHPGLIGLIGPPG---- 1423

Qy   253  RRRAKPSESRHPGP-GSRGRGGSIGL-GGGGMGPREAEPGELGIA-LRGGGAADPPFCP 309
               || ||       |    | | |       ||  ||     || |  | || ||     | |
Db  1424  EQGEKGDRGLPGPGPGSPGQKGEMGIPGASGPIGP-GGPGLPGPAGPRGKGATGPGP 1481

Qy   310  HYEK-VSGDYGHPVIVQDGPP 330
```

Db   1482 KGEKGVQGPPGHP------GPP 1497

Search completed: February 11, 2003, 12:03:44
Job time : 32.4591 secs

```
Db  165  PGAPGPPQPPAEKGAKGAMGRDGATGPSGPQGPPGVKGEAG------LQGPQGAPG-----  214
         |||                                                       |||
Qy  253  RRRAKPSESRHPGP-----GSFGRGGSLGLGGGGMGPREAEPGELGI------ALRGGG  301

Db  215  ----KCGATGTPGQGEKGSKGDGLIGPKGETG---TKGEKGDLGLPGSKGDRGMKGDA  267
             |||                                         ||
Qy  302  AADPFFCPHYEKVSGDYGHPVIVQDGPP  330
         |||                     |||
Db  268  GVMGP--PGAQGSKGDFGRP-----GPP  288
```

RESULT 50
US-09-978-192A-614
Sequence 614, Application US/09978192A
Patent No. US20020177553A1
GENERAL INFORMATION:
APPLICANT: Ashkenazi, Avi
APPLICANT: Baker, Kevin P.
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan
APPLICANT: Ferrara, Napoleon
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J
APPLICANT: Kljavin, Ivar J.
APPLICANT: Kuo, Sophia S.
APPLICANT: Napier, Mary A.
APPLICANT: Pan, James;
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Shelton, David L.
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
TITLE OF INVENTION: Acids Encoding the Same
FILE REFERENCE: P2630P1C9
CURRENT APPLICATION NUMBER: US/09/978,192A
CURRENT FILING DATE: 2001-10-15
PRIOR APPLICATION NUMBER: 09/918585
PRIOR FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/064249
PRIOR FILING DATE: 1997-11-03
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066364
PRIOR FILING DATE: 1997-11-21
PRIOR APPLICATION NUMBER: 60/077450
PRIOR FILING DATE: 1998-03-10
PRIOR APPLICATION NUMBER: 60/077632
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077641
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/076649
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077791
PRIOR FILING DATE: 1998-03-12
PRIOR APPLICATION NUMBER: 60/078004
PRIOR FILING DATE: 1998-03-13
PRIOR APPLICATION NUMBER: 60/078886
PRIOR FILING DATE: 1998-03-20

PRIOR APPLICATION NUMBER: 60/078936
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078939
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/079294
PRIOR FILING DATE: 1998-03-25
PRIOR APPLICATION NUMBER: 60/079656
PRIOR FILING DATE: 1998-03-26
PRIOR APPLICATION NUMBER: 60/079664
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079689
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079663
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079728
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079786
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079920
PRIOR FILING DATE: 1998-03-30
PRIOR APPLICATION NUMBER: 60/079923
PRIOR FILING DATE: 1998-03-30
PRIOR APPLICATION NUMBER: 60/080105
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080107
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080165
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080194
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080327
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080328
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080333
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080334
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/081070
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/081049
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081195
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081203
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081229
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081955
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081817
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081819
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081952
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081838
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/082568
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082569
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082704
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082804
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082700
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082797
```

```
Qy  198  PGKENLPGDPTSNATS-----RGAEGPLPPPSMPAVAGARAGGLALLLLGVAGAGGAMCWR 252
              |       |                          |   ||  |        ||||
Db  165  PGAPGPPGPPPAEKGAKGAMGRDGATGPSGGPQGPPGVKGEAG-----LQGPQGAPG----- 214

Qy  253  RRRAKPSESRHPGP-----GSFGRGGSLGLGGGGGMGPRREAEPGEIGI-----ALRGGG 301
                |       |     | |       :      | ||  ||||:
Db  215  ----KQGATGTPGPQGGEKGSSKGDGGLIGPXGETG---TKGEKGDLGLPGSKGDRGMKGDA 267

Qy  302  AADPPFCFHYEKVSGDYGHPVYIVQDGPP 330
                 :      :              |||
Db  268  GVMGP--PGAQGSKGDFGRP-----GPP 288
```

Search completed: February 11, 2003, 12:13:11
Job time : 25.4591 secs

~ PRIOR FILING DATE: 1998-04-22
~ PRIOR APPLICATION NUMBER: 60/082796
~ PRIOR FILING DATE: 1998-04-23
~ PRIOR APPLICATION NUMBER: 60/083336
~ PRIOR FILING DATE: 1998-04-27
~ PRIOR APPLICATION NUMBER: 60/083322
~ PRIOR FILING DATE: 1998-04-28
~ PRIOR APPLICATION NUMBER: 60/083392
~ PRIOR FILING DATE: 1998-04-29
~ PRIOR APPLICATION NUMBER: 60/083495
~ PRIOR FILING DATE: 1998-04-29
~ PRIOR APPLICATION NUMBER: 60/083496
~ PRIOR FILING DATE: 1998-04-29
~ PRIOR APPLICATION NUMBER: 60/083499
~ PRIOR FILING DATE: 1998-04-29
~ PRIOR APPLICATION NUMBER: 60/083500
~ PRIOR FILING DATE: 1998-04-29
~ PRIOR APPLICATION NUMBER: 60/083554
~ PRIOR FILING DATE: 1998-04-29
~ PRIOR APPLICATION NUMBER: 60/083559
~ PRIOR FILING DATE: 1998-04-29
~ PRIOR APPLICATION NUMBER: 60/083545
~ PRIOR FILING DATE: 1998-04-30
~ PRIOR APPLICATION NUMBER: 60/083558
~ PRIOR FILING DATE: 1998-04-29
~ PRIOR APPLICATION NUMBER: 60/083742
~ PRIOR FILING DATE: 1998-04-29
~ PRIOR APPLICATION NUMBER: 60/084414
~ PRIOR FILING DATE: 1998-04-30
~ PRIOR APPLICATION NUMBER: 60/084366
~ PRIOR FILING DATE: 1998-05-06
~ PRIOR APPLICATION NUMBER: 60/084441
~ PRIOR FILING DATE: 1998-05-07
~ PRIOR APPLICATION NUMBER: 60/084637
~ PRIOR FILING DATE: 1998-05-07
~ PRIOR APPLICATION NUMBER: 60/084639
~ PRIOR FILING DATE: 1998-05-07
~ PRIOR APPLICATION NUMBER: 60/084640
~ PRIOR FILING DATE: 1998-05-07
~ PRIOR APPLICATION NUMBER: 60/084598
~ PRIOR FILING DATE: 1998-05-07
~ PRIOR APPLICATION NUMBER: 60/084600
~ PRIOR FILING DATE: 1998-05-07
~ PRIOR APPLICATION NUMBER: 60/084627
~ PRIOR FILING DATE: 1998-05-07
~ PRIOR APPLICATION NUMBER: 60/084643
~ PRIOR FILING DATE: 1998-05-07
~ PRIOR APPLICATION NUMBER: 60/085339
~ PRIOR FILING DATE: 1998-05-13
~ PRIOR APPLICATION NUMBER: 60/085338
~ PRIOR FILING DATE: 1998-05-13
~ PRIOR APPLICATION NUMBER: 60/085323
~ PRIOR FILING DATE: 1998-05-13
~ PRIOR APPLICATION NUMBER: 60/085582
~ PRIOR FILING DATE: 1998-05-15
~ PRIOR APPLICATION NUMBER: 60/085700
~ PRIOR FILING DATE: 1998-05-15
~ PRIOR APPLICATION NUMBER: 60/085689
~ PRIOR FILING DATE: 1998-05-15
~ PRIOR APPLICATION NUMBER: 60/085579
~ PRIOR FILING DATE: 1998-05-15
~ PRIOR APPLICATION NUMBER: 60/085580
~ PRIOR FILING DATE: 1998-05-15
~ PRIOR APPLICATION NUMBER: 60/085573
~ PRIOR FILING DATE: 1998-05-15
~ PRIOR APPLICATION NUMBER: 60/085704
~ PRIOR FILING DATE: 1998-05-15
~ PRIOR APPLICATION NUMBER: 60/085697

Query Match   5.2%; Score 126.5; DB 9; Length 520;
Best Local Similarity 30.9%; Pred. No. 0.062;
Matches   46; Conservative   9; Mismatches 53; Indels 41; Gaps 8;

OM protein - protein search, using sw model

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: February 11, 2003, 11:47:56 ; Search time 33.7673 Seconds
                                              (without alignments)
                                              2776.398 Million cell updates/sec

Title: US-10-021-121-2
Perfect score: 2450
Sequence: 1 MGPPHSGPGGVRVGALLLG..........TTLLRQRASVEABAGQHGPL 455
Scoring table: BLOSUM62
                Gapop 10.0 , Gapext 0.5
Searched: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 50 summaries

Database :   SPTREMBL_21:*
                1: sp_archea:*
                2: sp_bacteria:*
                3: sp_fungi:*
                4: sp_human:*
                5: sp_invertebrate:*
                6: sp_mammal:*
                7: sp_mhc:*
                8: sp_organelle:*
                9: sp_phage:*
                10: sp_plant:*
                11: sp_rodent:*
                12: sp_virus:*
                13: sp_vertebrate:*
                14: sp_unclassified:*
                15: sp_rvirus:*
                16: sp_bacterial:*
                17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1550	63.3	285	4	Q8TBH7	Q8tbh7 homo sapien
2	914	37.3	331	13	Q90231	Q90231 brachydanio
3	910	37.1	327	13	Q9PT69	Q9pt69 xenopus lae
4	620.5	25.3	333	13	Q9PUJ4	Q9puj4 gallus gall
5	607	24.8	341	13	Q90233	Q90233 brachydanio
6	596.5	24.3	334	13	Q90232	Q90232 brachydanio
7	331	13.5	205	13	Q9W6H9	Q9w6h9 xenopus lae
8	214.5	8.8	237	13	Q9U3M2	Q9u3m2 caenorhabdi
9	198.5	8.1	279	5	Q9U474	Q9u474 caenorhabdi
10	192	7.8	652	5	Q9V4B1	Q9v4b1 drosophila
11	178.5	7.3	205	11	Q98TZ1	Q98tz1 gallus gall
12	167	6.8	205	11	Q9D7K8	Q9d7k8 mus musculu
13	163.5	6.7	206	11	Q9CZS8	Q9czs8 mus musculu
14	155.5	6.3	675	5	Q9N178	Q9n178 sus scrofa
15	155	6.3	229	13	Q93431	Q93431 brachydanio
16	153.5	6.3	1691	11	Q9ESQ2	Q9esq2 mus musculu

ALIGNMENTS

RESULT 1
Q8TBH7

ID	Q8TBH7;	PRELIMINARY;	PRT;	285 AA.	
AC	Q8TBH7;				
DT	01-JUN-2002 (TrEMBLrel. 21, Created)				
DT	01-JUN-2002 (TrEMBLrel. 21, Last sequence update)				
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)				
DE	Similar to ephrin B3 (Fragment).				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=BRAIN;				
RA	Strausberg R.;				
RL	Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.				
DR	EMBL; BC022499; AAH22499.1; -.				
FT	NON_TER 1				
SQ	SEQUENCE 285 AA; 30159 MW; 08DBACE714246982 CRC64;				

Query Match   63.3%; Score 1550; DB 4; Length 285;
Best Local Similarity   100.0%;
Matches 283; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   56 DRLDILCPRARPPGPHSSSPNYEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDLLRFTI 115
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 DRLDILCPRARPPGPHSSSPNYEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDLLRFTI 60

Qy   116 KFQEYSPNLWGHEFRSHHDYYIIATSDGTRGLESLQGGVCLTRGMKVLLRVGQSPRGGA 175
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 KFQEYSPNLWGHEFRSHHDYYIIATSDGTRGLESLQGGVCLTRGMKVLLRVGQSPRGGA 120

Qy   176 VPRKPVSEMPMERDRGAAHSLEPGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAAGL 235

```
Db   121 VPRKPVSEMPMERDRGAAHSLEPGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAGGL 180
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   236 ALLLLGVAGAGGAMCWRRRRAKPSESRHPGPGSSFGRGGGSLGLGGGGMGPREAEPGELGI 295

Db   181 ALLLLGVAGAGGAMCWRRRRAKPSESRHPGPGSSFGRGGGSLGLGGGGMGPREAEPGELGI 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   296 ALRGGGAAADPPPCPHYEKVSGDYGHPVYIVQDGPPQSPPNIYY 338

Db   241 ALRGGGAAADPPPCPHYEKVSGDYGHPVYIVQDGPPQSPPNIYY 283
```

RESULT 2

```
Q90Z31
ID   Q90Z31                         PRELIMINARY;      PRT;    331 AA.
AC   Q90Z31;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Ephrin B3.
OS   Brachydanio rerio (Zebrafish) (Zebra danio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21290827; PubMed=11397014;
RA   Chan J., Mably J.D., Serluca F.C., Chen J.N., Goldstein N.B.,
RA   Thomas M.C., Cleary J.A., Brennan C., Fishman M.C., Roberts T.M.;
RT   "Morphogenesis of prechordal plate and notochord requires intact
RT   eph/ephrin b signaling.";
RL   Dev. Biol. 234:470-482(2001).
DR   EMBL; AF375227; AAK64277.1; -.
DR   InterPro; IPR001799; Ephrin.
DR   Pfam; PF00812; Ephrin; 1.
DR   ProDom; PD002533; Ephrin; 1.
DR   PROSITE; PS01299; EPHRIN; UNKNOWN_1.
SQ   SEQUENCE  331 AA;  35638 MW;  6A5EACD509A09818 CRC64;

Query Match       37.3%;  Score 914;  DB 13;  Length 331;
Best Local Similarity   54.5%;  Pred. No. 6.4e-64;
Matches  188;  Conservative  43;  Mismatches  74;  Indels  40;  Gaps  10;

Qy    10 GVRVGALLLLGVLGLVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDLLCPRARPPG 69
             :||: ||:||||:|| :|  ||:  ||::  :::  : |: ||||||||||||||
Db     9 GLGILLFLVDLLG--ITATNMEPIYWNSLNKRFSDDKGYVLYPIGDRLDLICPSSDPPG 67

Qy    70 PHSSPNYEFYKLYLVGG-AQGRRCEAPPAPNLLLTCDRPDLDLRFTIKFQEYSPNLWGHE 128
             | ||:|:||||:||   :|| ||  || || :||||||||||| ||:||||||||||
Db    68 PRAPADVEYYKLYLVSSREQADRCEVTGAAPNLLLTCDKPNSDMRFTIKFQEYSPNLWGHE 127

Qy   129 FRSHHDYYIIAISDGTREGLESLQGGVCLTRGMKVLLRVGQSPRG-GAVPRKPVSEMPME 187
             |:||||||||::|||| || ||:||:|:::::||||:|:||| :|  :::::::|::::
Db   128 FKINHDYPIIAISDGTRQGLESMRGGVCATQGMKVVLKVGQSPYGLPAKSPKPDS----- 182

Qy   188 RDRGAAHSLEPGKENLPGDPTSNAT----------SRGAEGPLPPPSMPAVAGAAGGLALLL 239
             : |    :|:||:||:: ||:|::|         : :::::|||:::: :||::::::|||
Db   183 ------AGRINNPFGTGNSTHPQIPRGSGGENGPLPASNIAVIAGAAGGSAFLL 232

Qy   240 LGVAGAGGAMCWRRRRAKPSESRHPGPGSFG-----RGGSLGLGGGGMGPREAEPGEL 293
             ||:|| ::::::::::|:::|: |:|::|:    |::||:||| ::::::::|:
Db   233 L-VTAVICVVCYRRRHAKHSESHHP-PLSLSSLTSPKRGCGCGGGGNNNG----SEPSDI 287

Qy   294 GIALRGGGAAADPPPCPHYEKVSGDYGHPVYIVQDGPPQSPPNIYY 338
             ||::|:|::|||||:|:||||||||||||:||:|: :|:|:::::
Db   288 IIPLR----TSDSAYCPHYEKVSGDYGHPVYIVQEMPQSPANIYY 329
```

RESULT 3

```
Q9PT69
ID   Q9PT69                         PRELIMINARY;      PRT;    327 AA.
AC   Q9PT69;
```

```
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Ephrin-B3 precursor.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=EMBRYO;
RX   MEDLINE=20099673; PubMed=10633856;
RA   Helbling P.M., Saulnier D.M.E., Robinson V., Christiansen J.H.,
RA   Wilkinson D.G., Brandli A.W.;
RT   "Comparative analysis of embryonic gene expression defines potential
RT   interaction sites for Xenopus EphB4 receptors with ephrin-B ligands.";
RL   Dev. Dyn. 216:361-373(1999).
DR   EMBL; AJ236866; CAB65511.1; -.
DR   InterPro; IPR002088; Aldehyde_dehydr.
DR   InterPro; IPR001799; Ephrin.
DR   Pfam; PF00812; Ephrin; 1.
DR   ProDom; PD002533; Ephrin; 1.
DR   PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; UNKNOWN_1.
DR   PROSITE; PS01299; EPHRIN; 1.
KW   Signal.
FT   SIGNAL        1      24       POTENTIAL.
SQ   SEQUENCE  327 AA;  35913 MW;  4BB0FA39D4C22DCD CRC64;

Query Match       37.1%;  Score 910;  DB 13;  Length 327;
Best Local Similarity   60.1%;  Pred. No. 1.3e-63;
Matches  190;  Conservative  30;  Mismatches  82;  Indels  14;  Gaps  8;

Qy    25 VSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDLLCPRARPPGPHSSPNYEFYKLYLV 84
             :|:|||:||:|||||||||:|||||||| :|||||||||||||:|| :::||:|||||||:
Db    22 ISALSLDPIIWNSSNKRFEDTEGYVLYPQIGDRLDLLCPRSEPQGPFSSSSPYEYKLYLIV 81

Qy    85 GGAGQ-RRCEAPPAPNLLLTCDRPDLDLRFTIKFQEYSPNLWGHEFRSHHDYYIIAISDG 143
             |||:| :|||:||:|||||||||| :|:|||:|||||||||||||||::|||||||::|||
Db    82 GTKEEMSSCSILRTPNLLLTCDRPSQDLRFTIKFQEFSPNLWGHEFQOSRDYYIIATSDG 141

Qy   144 TREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKPVSEMPMERDRGAAHSL-EPGKEN 202
             |:|||:||||||:::|||||||:||||||| :::::|:|::::||||||||:|  |||:|
Db   142 TMDGIETLCGGVCETKGMKVTLKVGQSPNGATPPRRPSS---AGKDSGTSPSVPNPDIPN 198

Qy   203 LPGDPTSNATSRGAEGPLPPSMPAVAGAAGGLALLLLGVAGAGGAMCWRRRRAKPSESR 262
             ||||||||:|:|:||||||::||||||||||||:|||::|||||||||||||||||:::|
Db   199 V-GETEGNATKTGENGPLPISHVPLVAGAAGGAALLLL-VFGVVGWVCHRRRAQKHSDTR 256

Qy   263 HPGPGSFGRGGGSLGLGGGGMGPREAEPGELGIALRGGGAAADPPPCPHYEKVSGDYGHPV 322
             ||||||| :::::|:| ::|:|||::: |:| ||:|:|:::|||||||||||||||||||||
Db   257 HP-FLSLGSITSPKRGGNN----NGHEPSDIIMPLRPSEAG--AFCPHYEKVSGDYGHPV 309

Qy   323 YIVQDGPPQSPPNIYY 338
             |||:: :|:|::|||
Db   310 YIVQDMASQSPANIYY 325
```

RESULT 4

```
Q9PUJ4
ID   Q9PUJ4                         PRELIMINARY;      PRT;    333 AA.
AC   Q9PUJ4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Ephrin-B2 precursor.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RA   Menzel P., Pasquale E.B.;
RT   "Coding sequence of chicken ephrin-B2.";
RL   Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF180729; AAD53948.1; -.
DR   InterPro; IPR001799; Ephrin.
DR   Pfam; PF00812; Ephrin; 1.
DR   ProDom; PD002533; Ephrin; 1.
DR   PROSITE; PS01299; EPHRIN; 1.
KW   Signal.
FT   SIGNAL          1     27
FT   CHAIN          28    333        EPHRIN-B2.
FT   SIGNAL          1     27        POTENTIAL.
SQ   SEQUENCE   333 AA;  36761 MW;  4C28E8C821187783 CRC64;

Query Match             25.3%;  Score 620.5;  DB 13;  Length 333;
Best Local Similarity   40.2%;  Pred. No. 7.2e-41;
Matches  133;  Conservative  52;  Mismatches 125;  Indels  21;  Gaps  6;

Qy    14 GALLLLGLVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDLLCPRARPPGPHSS    73
            : || :: |:|||||:||||||||:||||  |:| |:|:| ::||||:..:
Db    16 GALMVLWRTALAKSIVLDP-YWNSSNPKFLPGGQVLYPQIGDKLDIICPKV---DSKTV  72

Qy    74 PNYEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDLDLRFTIKFQEYSPNLWGHEFRSHH  133
            :|| :|  :|  |  |.:|||||||||||||||||||||||||||:||||||..:
Db    73 GQYEYKVXVWDKQQADSCAIRKDNTPLLNCAKPDQDVKFTIKFQEFSPNLWGLEFQKNK  132

Qy   134 DYYIIATSDGTREGLESLQGGVCLITRGMKVLLRVGQSPRGGAVPRKPVSEMPMER----  188
            |||:::||:| :||| | :|||| .:.||:|::|:|| |.     ||..
Db   133 DYYVISTSNGSLEGLNNQEGGVCQTKTMKILMKVGQDPNADSGALPR---STDPTRPEQEA 189

Qy   190 GTNGKSSTTSPFVKDHSGSSTDG--SKAGHSSILGSEVALFAGIASGCIFIVIILVV     247
            |.: .:. . :: : .:.    |||||||||:::.:||||.|||:||::::|:|::
Db   189 -DRGAAHSLEPGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGNAGGLALLLLGVAGAGG  247

Qy   248 AMCWRRRRAKPSESRHPGPGSFGRGGSLGLGGGGMGPREAEPGELGIALRGGGAADPPF  307
            .:||||||:::.. ... . ::.:... :  .: .:.    ::   :|.||.. |..
Db   248 LLLKYRRHRHSFQQHTTLSLSTLATPKRSGNN----NGSEPSDIIIPLR---TADSVF   300

Qy   308 CPHYEKVSGDYGHGHPVYIVQDGPPQSPPNIIY  338
            |||||||||||| |||||||::.||||||:|||
Db   301 CPHYEKVSGDYGHPVYIVQEMPPQSPANIIY    331


RESULT 5
Q90233
Q90233
ID   Q90233;  PRELIMINARY;  PRT;  341 AA.
AC   Q90233;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Ephrin B1.
OS   Brachydanio rerio (Zebrafish) (Zebra danio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21290827; PubMed=11397014;
RA   Chan J., Mably J.D., Serluca F.C., Chen J.N., Goldstein N.B.,
RA   Thomas M.C., Cleary J.A., Brennan C., Fishman M.C., Roberts T.M.;
RT   "Morphogenesis of prechordal plate and notochord requires intact
RT   eph/ephrin b signaling";
RL   Dev. Biol. 234:470-482(2001).
DR   EMBL; AF375224; AAK64274.1; -.
DR   InterPro; IPR001799; Ephrin.
DR   Pfam; PF00812; Ephrin; 1.
DR   ProDom; PD002533; Ephrin; 1.
DR   PROSITE; PS01299; EPHRIN; 1.
SQ   SEQUENCE   341 AA;  37849 MW;  CB922F20E0D93B94 CRC64;

Query Match             24.8%;  Score 607;  DB 13;  Length 341;
Best Local Similarity   41.5%;  Pred. No. 8.5e-40;
Matches  139;  Conservative  49;  Mismatches 109;  Indels  38;  Gaps  11;
```

```
Qy    24 LVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDLLCPRARPPGPHSSSPNYEFYKYL   83
            | ..:|||.:|||:|:.||:...|.|:||:|||:||::||| .     |.||:|| ||||
Db    23 LPAAKSLESVVWNSQNPKFVSGKGLVIYPEIGDKLDIICPK-----GDMGRP-YEFYKLYL  77

Qy    84 VGGAQGRRCEAPPAPNLLLTCDRPDLDLRFTIKFQEYSPNLWGHEFRSHHDYYIIATSDG  143
            |.. :. ||.:.::.|:| ||.:|: ::|||||||:|||||||:||| .:.|||::::.|
Db    78 VKKEQAESCCSTILDPNVLVTCNKPEKDIKFTIKFQEFSPNLWGLEFRFTNYYITSTSNG  137

Qy   144 TRGGLESLQGGVCLITRGMKVLLRVGQSPRG-----GAVPRKPVSEMPMERDRGAAHSLEP 198
            |:.|||.::|||| .|| |:|::::|||.|..     |.:   :  ...:|  ..
Db   138 TQEGLENREGGVCGSTRSMKIIMKVGQDPNADDGDLPDLPDRYDNEIKDPTTSPSRKTER  197

Qy   199 GKEN-------------LPGDPTSNATSR--GAEGPLPPPSMPAVAGAAG-GLALLLLGVAGAGG 247
            |:||             || : . . ..   |:.. ... ::||::|:|. |...:::.:|.:.:
Db   198 GRENEVDGNGSRKMPGKDTRNQNNSPGSVEGIP--GSKPALFAAIGAGCVIFLLIIIIV    255

Qy   248 AMCWRRRRAKPSESRHPGPGSFGRGGSL----GLGGGGMGPREAEPGELGIALRGGGAA  303
            .:.: :: :.. ..  . . ::. ...:     :  . .:.   .:.|.| :.||.:..
Db   256 LLLKLRKITR-----KHSQP----RGGTALSLSTLATPKGAAQAGSEPSDIIIPLR---TT  304

Qy   304 DPPFCPHYEKVSGDYGHGHPVYIVQDGPPQSPPNIIY 338
            ...:|||||||||||| ||||||||::.||||||:|||
Db   305 ENNYCPHYEKVSGDYGHPVYIVQEMPPQSPANIIY    339


RESULT 6
Q90232
Q90232
ID   Q90232;  PRELIMINARY;  PRT;  334 AA.
AC   Q90232;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Ephrin B2b.
OS   Brachydanio rerio (Zebrafish) (Zebra danio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21290827; PubMed=11397014;
RA   Chan J., Mably J.D., Serluca F.C., Chen J.N., Goldstein N.B.,
RA   Thomas M.C., Cleary J.A., Brennan C., Fishman M.C., Roberts T.M.;
RT   "Morphogenesis of prechordal plate and notochord requires intact
RT   eph/ephrin b signaling";
RL   Dev. Biol. 234:470-482(2001).
DR   EMBL; AF375226; AAK64276.1; -.
DR   InterPro; IPR001799; Ephrin.
DR   Pfam; PF00812; Ephrin; 1.
DR   ProDom; PD002533; Ephrin; 1.
SQ   SEQUENCE   334 AA;  36998 MW;  34149720FE9473BD CRC64;

Query Match             24.3%;  Score 596.5;  DB 13;  Length 334;
Best Local Similarity   40.6%;  Pred. No. 5.5e-39;
Matches  132;  Conservative  45;  Mismatches 113;  Indels  35;  Gaps  8;

Qy    30 LEPVYWNSANKRFQAEGGYVLYPQIGDRLDLLCPRARPPGPHSSSPNYEFYKYLVLGGAQG  89
            | .:|||:.|.||.. . |.|||.|||:.|::||| : ||.:.:  |.|| |.||| .:
Db    27 LSIYWNTSNTKFVPGRGVVLYPIGDKMDIVCPRIK-PGSTEQTNIEYFRYVLVPKEQL    85

Qy    90 RRCEAPPAPNLLLTCDRPDLDLRFTIKFQEYSPNLWGHEFRSHHDYYIIATSDGTREGLE  149
            :.| ..::.|:|| ||:||: ||:|||:|||:|||||||:||| .||:|:::::.|:||.
Db    86 ETCHVTKSDMLLLNCRKPDQDVKFTFKFQEFSPNLWGLEFLRGKDYHIISTSNSTFEGLD  145

Qy   150 SLQGGVCLITRGMKVLLRVGQSPRGGAV----PRKPVSEMPMERDRGAAHSLEPGKENLPG 205
            .::|||| .. ||::::|||||| .. |    |::    . .:. .. ..  .|:
Db   146 NHHGGVCRSKSMKLIVLRVGQSPTDSFSAKNHPTRNPKYPENKDQNTF----SKENDVS  200

Qy   206 --DPTSNATSRGAEGPLPPPSMPAVAGAGAAGGIALLLLGVAGAGGAMCW---------- 253
             . . ..::...:..::. . ::||.|..:..:| :::..:..|.:.
Db   201 QIDSMQGBSGGKSG---------BSVGSAGSDVALFAGV--ASGAVIFILIIIALVALLH  250
```

```
Qy   254 RRAKPSESRHPGPCSFGRGGSLGLGGGGGMGPREAPEGELGIALRGGGAADPPFCPHYEK 313
             ||||  |:  |:||  |  :|||||  ||           ||
Db   251 RRHQKHSAQCSGQLPLNTLPKRGSGASGGSNNNGSEPSDIIFPIRTSGSM---YCPHYEK 307

Qy   314 VSGDYGHPVYIVQDGPPQSPPNIYY 338
             |||||||||||||:|  || ||||
Db   308 VSGDYGHPVYIVQEMPPQNPANIYY 332


RESULT 7
Q9W6H9
ID   Q9W6H9    PRELIMINARY;    PRT;    205 AA.
AC   Q9W6H9;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Ephrin-B2 (Fragment).
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97411149; PubMed=9259557;
RA   Smith A., Robinson V., Patel K., Wilkinson D.G.;
RT   "The EphA4 and EphB1 receptor tyrosine kinases and ephrin-B2 ligand
RT   regulate targeted migration of branchial neural crest cells.";
RL   Curr. Biol. 7:561-570(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=2009673;
RA   Helbling P.M., Saulnier D.M., Robinson V., Christiansen J.H.,
RA   Wilkinson D.G., Brandli A.W.;
RT   "Comparative analysis of embryonic gene expression defines potential
RT   interaction sites for Xenopus EphB4 receptors with ephrin-B ligands.";
RL   Dev. Dyn. 216:361-373(1999).
DR   EMBL; AF128844; AAD32610.1; -.
DR   InterPro; IPR001799; Ephrin.
DR   Pfam; PF00812; Ephrin; 1.
DR   ProDom; PD002533; Ephrin; 1.
FT   NON_TER    1    1
SQ   SEQUENCE    205 AA;    22256 MW;    7DEDD34277260C87 CRC64;

Query Match         13.5%;     Score 331;    DB 13;    Length 205;
Best Local Similarity    36.4%;     Pred. No. 2.2e-18;
Matches    82;    Conservative    35;    Mismatches    72;    Indels    36;    Gaps    7;

Qy   128 EFRSHHDYYIIATSDGTREGLESLQGGVYCLTRGMKVLLRVGQSP----RGGAVPRKPVS 182
             |:  ||||| |||||||||||||      ||||||        |:|
Db   1   BFQRDKDYYIISTSNGSLEGVDNQBEGGVCVTKAMKILMKVGQDPNFENHRGASSTRPDH 60

Qy   183 EMPM--ERDRGAAHSLEPGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAGGLALLLL 240
             |:|   || ||| |
Db   61  ESGTNGKSSTTSPHVNGBPEGSSTEGKNAGHSSIIGSEVAL-------PAGIASGSIFIV 113

Qy   241 GVAGAGGAWCWRRRAKSESRHPGPGSFGRGGSLGLG------------GGGGMGPREAPEGEL 293
             :|||
Db   114 IIITLVVLLLKYRHRHRHSPQHT------------TLSLSTLATPKRSGNNNG---SEPSDI 161

Qy   294 GIALRGGGAADPPPCPHYEKVSGDYGHPVYIVQDGPPQSPPNIYY 338
             |:          |  |||||||||||||||:|||| ||||
Db   162 IIPLR---TAEGVPCPHYEKVSGDYGHPVYIVQEMPPQQNPANIYY 203


RESULT 8
Q9U3M2
ID   Q9U3M2    PRELIMINARY;    PRT;    237 AA.
AC   Q9U3M2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   C43F9.8 protein.
GN   C43F9.8.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Mortimore B.J.;
RL   Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   none;
RX   MEDLINE=99069613; PubMed=9851916;
RT   "Genome sequence of the nematode C.elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
DR   EMBL; Z82262; CAB54195.1; -.
DR   InterPro; IPR001799; Ephrin.
DR   Pfam; PF00812; Ephrin; 1.
DR   ProDom; PD002533; Ephrin; 1.
SQ   SEQUENCE    237 AA;    26748 MW;    B9B2D9FCC71FE4FC CRC64;

Query Match          8.8%;    Score 214.5;    DB 5;    Length 237;
Best Local Similarity    25.7%;    Pred. No. 3.7e-09;
Matches    53;    Conservative    41;    Mismatches    83;    Indels    29;    Gaps    6;

Qy   11  VRVGALLLLGVLGLVS-GLSLEPVTWNSANKRFQAEG-GVVLYFQIGRDLDLLCPRAPP 68
             |::||| |||||      |    |:
Db   1   MQIATILLLSLPFFIGWARKIPDINWISSNPIFDVSNTDHVISVHIIGDRVSIRCPKSDET 60

Qy   69  GPHSPNYEFYKLYLYVGGAGGRRCRAPPAPNLLIICDRPDLDLRFTIKFQEYSPNLWGHE 128
             |    ::|  |:|
Db   61  G-----KYEYSYIYMVSDEEYDHCFL-SKPRLVGACDNQTINASINIVFRSFTPTPGGFE 114

Qy   129 FRSHHDYYIIA------------------TSDGTREGLESLQGGVYCLTRGMKVLLRVGQ 169
             | :  |:||       |||||||||||||     |||:||||
Db   115 FQPGKNYFLISKSEVDALIIYETANQIFPGTSDGTLEGIDRKDGLCTAKQMKIKFEVGQ 174

Qy   170 SPRGGAVPRKPVSVSEMPMERDRGAAHS 195
             |:  :   ||         |||| |||
Db   175 DRRGIENPK--PAARTLKKQRDAEHS 198


RESULT 9
Q9U474
ID   Q9U474    PRELIMINARY;    PRT;    279 AA.
AC   Q9U474;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   VAB-2 (Hypothetical protein Y37E11AR.6).
GN   VAB-2 OR Y37E11AR.6.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=N2;
RX   MEDLINE=20084449; PubMed=10619431;
RA   Chin-Sang I.D., George S.E., Ding M., Moseley S.L., Lynch A.S.,
RA   Chisholm A.D.;
RT   "The ephrin VAB-2/EFN-1 functions in neuronal signaling to regulate
RT   epidermal morphogenesis in C. elegans.";
RL   Cell 99:781-790(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   none;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
```

```
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Miller N., Maggi L.;
RL   "The sequence of C. elegans cosmid Y37E11AR.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RT   "Direct Submission.";
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF201079; AAK25647.1; -.
DR   EMBL; AC024759; AAK68436.1; -.
DR   InterPro; IPR001799; Ephrin.
DR   Pfam; PF00812; Ephrin; 1.
DR   ProDom; PD002533; Ephrin; 1.
SQ   SEQUENCE 279 AA; 32068 MW; 8C291A92D97D39EF CRC64;

Query Match   8.1%; Score 198.5; DB 5; Length 279;
Best Local Similarity 27.8%; Conservative 35; Mismatches 90; Indels 41; Gaps 8;
Matches 64;

Qy     1 MGPPHSGPGGVRVGAALLLLGVLGLV--SGLSLSEPVYWNSANKRFQAEGGYVLYPQIGDRL 58
Db     1 MHPP------IKIQTILJF-ILTTVHCSAXRLPQIYWNSTNPLVER------YAAIGDTL 47

Qy    59 DLLCPRARPPGPHSSPNYEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDLDLRTIKFPQ 118
Db    48 DIVCPFF---DENSDELIEQSIIYRVTEEEYENCERRSKAKELGRCTQPYQSEKLKVAFR 104

Qy   119 EYSPNLWGHEFRSHHDYIIIATSDGTRGELESLQGGVCLTRGMKVLLRVGQSPRGGAVPR 178
Db   105 LMSPNPSGLDYRFGVTYTYFISTSTGSRKGLYNEQCGGLCASHNLKMVIHI--TDRNG-- 158

Qy   179 KPVSEMPMERDRGAAHSLEPGKENLPGDPTSNATSRGAEGPLPPPSMPEAV 228
Db   159 -------DIGPHHHRHHHKKTTTTTTSTSTSTS-------TPKTIPPV 191
```

```
RESULT 10
ID   Q9V4E1      PRELIMINARY;      PRT;   652 AA.
AC   Q9V4E1;
DT   01-MAY-2000 (TREMBLrel. 13, Created)
DT   01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT   01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE   Ephrin protein.
GN   EPHRIN OR CG1862.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Murray D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Dai Y., Kunes S.;
RT   "Isolation and Characterization of Drosophila Ephrin.";
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF212287; AAF28394.1; -.
DR   EMBL; AE003843; AAF59335.2; -.
DR   FlyBase; FBgn00403324; Ephrin.
DR   InterPro; IPR001799; Ephrin.
DR   Pfam; PF00812; Ephrin; 1.
DR   ProDom; PD002533; Ephrin; 1.
SQ   SEQUENCE 652 AA; 72302 MW; 5BA2F02F15964594 CRC64;

Query Match   7.8%; Score 192; DB 5; Length 652;
Best Local Similarity 21.2%; Conservative 64; Mismatches 154; Indels 164; Gaps 22;
Matches 103;

Qy    33 VYWNSANKRFQAEG-GYVLYPQIG-----DRLDLLCPRARPPGPHSSPNYEFYKLYLVG 85
Db   219 MHWNTSNSIFRIDNTDHIIDVNKGNLAFEFDQVHIICP-VYEPGTFENET-EKYIIYNVS 276

Qy    86 GAQGRRCEAPPA-PNLLLTCDRPDLDLRFTIKFQEYSPNLWGHEFRSHHDYIIIATSDGT 144
Db   277 KVEYETCRITNADPRVIAICDKPQKLMFFITTRRPTTQPGGLEFLPGNDYYFISTS--S 334

Qy   145 REGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKPVSEMPMERDRGAAHSL-----EPG 199
Db   335 KDDLYRIGGRGSTNNMKVVPKVCCAPEDNN-KTTALSNSKSVTDTGGAINVNIANNDES 393

Qy   200 KENLPGDPTSNATSRGAEG-------------PLPP---------------------P 223
Db   394 HVNSHGNNIA.IGTNIGINGGQIIGGPQOSAGTPINFLSGNNNINGIPTTINSNIDQPNRIP 453

Qy   224 SMPAVAGAAGGLALLLLGVAGAGGAMCWRRRRAKFSBSRHPGPG-SFGRGGSIGLGGGGG 282
Db   454 IQPNIIGNHVGTNAVGTGIVGGGGIIL-------TPGHAHGNIMLQPGRGGI 499

Qy   283 MGPREAEPG---ELGIALRG------------GGAADDPPFCPHYEKVSGDYGH 320
Db   500 NG----AYPGHHHIQTGIRINNVPTQEHYPSHKGNANSNINGNDDH--HHYNK---H 547

Qy   321 PVYIVQD------GPFQSPPNIYYTSISVLBWPILHTIQLFFMSRKCSRVTTFLFPVQV 373
Db   548 PNEVVKNEELIYNSGAATSDGNIFALWWILS----------IFP--L 583

Qy   374 ITTTSTCRMTSFSFTTLNPSMQACRAQMGEFRIRWCFWGDRILGTALFVLVLILLGRLNM 433
Db   584 LSIQSCHLSSY------------WI-------SASFLVSTIAILGIHYL 613

Qy   434 HQTTL 438
```

RESULT 11
Q98TZ1

ID   Q98TZ1   PRELIMINARY;   PRT;   202 AA.
AC   Q98TZ1;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   Ephrin-A6 (Fragment).
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Menzel P., Valencia F., Godement P., Dodelet V.C., Pasquale E.B.;
RT   "Ephrin-A6, a new ligand for EphA receptors in the developing visual
RT   system.";
RL   Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF312286; AAK00944.1; -.
DR   InterPro; IPR001799; Ephrin.
DR   Pfam; PF00812; Ephrin; 1.
DR   PRINTS; PR01347; EPHRIN.
DR   ProDom; PD002533; Ephrin; 1.
DR   PROSITE; PS01299; EPHRIN; 1.
FT   NON_TER   1   1
SQ   SEQUENCE   202 AA; 22624 MW; 91E2716FF9135379 CRC64;

Query Match   27.5%; Score 178.5; DB 13; Length 202;
Best Local Similarity 27.3%; Pred. No. 2.1e-06;
Matches 56; Conservative 27; Mismatches 84; Indels 37; Gaps 7;

Qy   33 VYWNSANKRFQAEGGYVLYPQIGDRLDLLCPRARPPGPHSSPNYEFYKLVLVGGAQGRRC 92
              ||||||| |: :||:| :::||| |||| ||||::
Db   25 VYWNGSNPRF-LQDDYSIQVSINDHLDIYCPHYSAPTPWA-----ESFTLFMVDEEGYRGC 79

Qy   93 EAPPAPNLLLTCDR----PDLDLRFTIKFQEYSPNLWGHEFRSHHDYIIIAT-SDGTREGL 148
              :|| :: : ||: |: | | ::|: :: ||
Db   80 SETPGAPKRWEGNKFPAPFVPVRFSEKIQRTFFSLGFEFRPGETYYISVPTPGS---- 135

Qy   149 ESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKPVSEMPMERDRGAAHSLEPGKENLPGDPT 208
              |:| |: ||: |: | :::::::
Db   136 ----AGRCLKLRVSVCCR----------ASTPEPLTEVPNSQPRGR----------GGPE 171

Qy   209 SNATSRGAEGPLPPPSMPAVAGA 232
              : ::
Db   172 GDAGSPRDAAPIPQRSRTLVALA 195

RESULT 12
Q9D7K8

ID   Q9D7K8   PRELIMINARY;   PRT;   205 AA.
AC   Q9D7K8;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Adult male tongue cDNA, RIKEN full-length enriched library,
DE   clone:2310004J15, full insert sequence.
GN   EFNA1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=TONGUE;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AK009144; BAB26102.1; -.
MGD;   MGI:103226; Efna1.
DR   InterPro; IPR001799; Ephrin.
DR   InterPro; IPR001680; WD40.
DR   Pfam; PF00812; Ephrin; 1.
DR   PRINTS; PR01347; EPHRIN.
DR   ProDom; PD002533; Ephrin; 1.
DR   PROSITE; PS01299; EPHRIN; 1.
DR   PROSITE; PS00678; WD_REPEATS_1; UNKNOWN 1.
SQ   SEQUENCE   205 AA; 23772 MW; E37E55767459A4EC CRC64;

Query Match   43%; Score 167; DB 11; Length 205;
Best Local Similarity 26.1%; Pred. No. 1.7e-05;
Matches 43; Conservative 34; Mismatches 76; Indels 12; Gaps 4;

Qy   18 IIGVLGLIVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDLLCPRARPPGPHSSPNYE 77
              |:| || | | ::||||| ::::| ||| ||||::| ||| |||| |||||:::||:|
Db   8   LIGLCCSLAAADRHIVFWNSSNPKFREE-DYTVHVQLNDYLDIICPHYEDDSV-ADAAME 65

Qy   78 FYKLVLVGGAQGRRCEAPPAPNLLLTCDRPDL---DLRFTIKFQEYSPNLWGHEFRSHHD 134
              || :| |: || :|| | :: ||| || :|| : || | ||: |
Db   66 RYTLYMVEHEYVACQPQSKDQVRWCNRPSAKHGPEKLSVKPQRFTPFILGKEFKEGHS 125

Qy   135 YYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRK 179
              || :: :|| : :::| || ||| | |:: || ::
Db   126 YYYISKPIYHQB--------SQCLKLKVIVNGKITHNPQAHVNPQE 163

RESULT 13
Q9CZS8

ID   Q9CZS8   PRELIMINARY;   PRT;   206 AA.
AC   Q9CZS8;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   10 days embryo cDNA, RIKEN full-length enriched library,
DE   clone:2610529M21, full insert sequence.
GN   EFNA4.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=EMBRYO;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
,

RA   Blake J., Boffelli D., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AK012195; BAB28092.1; -.
DR   MGD; MGI:106643; Efna4.
DR   InterPro; IPR001799; Ephrin.
DR   Pfam; PF00812; Ephrin; 1.
DR   PRINTS; PR01347; Ephrin.
DR   ProDom; PD002533; Ephrin; 1.
DR   PROSITE; PS01299; EPHRIN; 1.
SQ   SEQUENCE   206 AA; 22859 MW; 675E32971D1C6EBC CRC64;

Query Match   6.7%; Score 163.5; DB 11; Length 206;
Best Local Similarity   28.1%;
Matches   61; Conservative   17; Pred. No. 3.2e-05;
Mismatches 80; Indels 59; Gaps 10;

```
Qy   32 PVYWNSANKRFQAEGGYVLYFQIGDRLDLLCPRARPPGPHSSPNYEFYKLYLVGGAQGRR 91
            :|||:||| |:   |  :|||:||||:||  :|:::::| |:|: :|  ||:::  :  :|
Db   29 PIYWNSSNPRL-LRGDAVVELGFNDYLDIFCPHYESPGPPEGP--ETFALYIVDWSGYEA 85

Qy   92 CEAPPAPNL-LLTCDRPDL--DLRFTIKFQEYSPNLWGHEFRSHHDYYIIATSDGTREG 147
Db   86 CTAEGANSFQRWNCSMFPAPFSPVRFSEKIQRYTPFPLPGEFTYYISVPTPESPG 145

Qy  148 -LESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKPVSEMPMERDRGAAH-SLEPGKENLPG 205
            ::  :  |::|                              :: :::  :::
Db  146 RCLRLQVSVCC----------------------------KESGSSHESAHP-----VG 170

Qy  206 DPTSNATS--RGAEGPLPPPSMPAVAGAAGGLALLLL 240
            |     |                     ---:LCLLLL 194
Db  171 SPGESGTSGWRGGHAPSP---------LCLLLL 194
```

RESULT 14
ID   Q9N178
AC   Q9N178;   PRELIMINARY;   PRT;   675 AA.
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Type X collagen.
GN   COL10A1.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21015405; PubMed=11310976;
RA   Nielsen V.H., Bendixen C., Arnbjerg J., Sorensen C.M., Jensen H.E.,
RA   Shukri N.M., Thomsen B.;
RT   "Abnormal growth plate function in pigs carrying a dominant mutation
RT   in type X collagen.";
RL   Mamm. Genome 11:1087-1092(2000).
DR   EMBL; AF222861; AAF37271.1; -.
DR   InterPro; IPR001073; C1q.
DR   InterPro; IPR000087; Collagen.
DR   Pfam; PF00386; C1q; 1.
DR   Pfam; PF01391; Collagen; 1.
DR   PRINTS; PR00007; COMPLEMNTC1Q.
DR   ProDom; PD000007; Collagen; 2.
DR   SMART; SM00110; C1Q; 1.
DR   PROSITE; PS01113; C1Q; 1.
KW   Collagen.
SQ   SEQUENCE   675 AA; 65447 MW; 26397B103103B3F9 CRC64;

Query Match   6.3%; Score 155.5; DB 6; Length 675;
Best Local Similarity   28.5%; Pred. No. 0.00053;
Matches   103; Conservative   17; Mismatches 113; Indels 129; Gaps 25;

```
Qy    1 MGPPHSGPEGPGVRVGALILLGVILGLVSGLSLEPVYWNSANKRFQAEGGYVLYPQ-GDR-L 58
Db  212 MGPP--GPPGV-----------------------------GKR--GENGPPGGPGIKGDRGF 240

Qy   59 DLLCPRARPPGPHSSPNYEFYKLYLVGGAQGRR-CEAPPAPNLLLTCDRPDLDLRFTIKF 117
Db  241 PGEGSGPAGPPGPQGPP------------GEQGREGIGKPGAPG--AAQGPGL------ 278

Qy  118 QEYSPNLWGHEFRSHHDYYIIATSDGT----REGLESLQGGVCLTRGMKVLLRVGQSPRG 173
Db  279 ----PGTKGHPGAPG------MAGPPGAPGPGKPGLPGLKG----QRG-----------PIG 315

Qy  174 GAVPRKPVSEMPMERDRG-AAHSLEPGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGA 232
Db  316 --LPGAPGA-----KGEQGPAGHPGEPGLTGPPG------SRGPGGPKGIPGNNGVPGPK 362

Qy  233 GGLALLLLGVIAGAGGAMCWRRR----AKPSESRIPG--------PGSFGRGGSLGLGGGGG 282
Db  363 GEIG--LAGPAGPFGCAKGERGPSGLDGKPGCYPGEPGPGLNGGPKGNPGLPGKGDPGIGGPPG 420

Qy  283 M-----GPREAE--PGELGIIA-LRGG------GAADPPPCPHYEKVSGDYGHPVYIVQDG 328
Db  421 LPGPVGPAGPAGAKGVPGHNGEAGPRGAPGIPGTRGPIGPGPIGPGFPGGSKGDPGNP-----G 474

Qy  329 PP 330
Db  475 PP 476
```

RESULT 15
ID   O93431
AC   O93431;   PRELIMINARY;   PRT;   229 AA.
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Ephrin A-L1.
OS   Brachydanio rerio (Zebrafish) (Zebra danio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Durbin L., Brennan C.H., Shiomi K., Cooke J.;
RT   "Eph signalling is required for segmentation and differentiation of
RT   the somites.";
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ006838; CAA07264.1; -.
DR   InterPro; IPR001799; Ephrin.
DR   InterPro; IPR003006; IG_MHC.
DR   Pfam; PF00812; Ephrin; 1.
DR   PRINTS; PR01347; Ephrin.
DR   ProDom; PD002533; Ephrin; 1.
DR   PROSITE; PS01299; EPHRIN; 1.
DR   PROSITE; PS00290; IG_MHC; UNKNOWN 1.
SQ   SEQUENCE   229 AA; 26615 MW; 86844G2F67AF6F5C CRC64;

Query Match   6.3%; Score 155; DB 13; Length 229;
Best Local Similarity   27.2%; Pred. No. 0.00017;
Matches   56; Conservative   23; Mismatches 93; Indels 34; Gaps 7;

```
Qy   33 VYWNSANKRFQAEGGYVLYPQIGDRLDLLCPRARPPGPHSSPNYEFYKLYLVGGAQGRRC 92
Db   25 VYWNSTNANFLWD-DYTVDVRINDYLDIICPH-YAHGEIASQEAERVYLYMVELEDYENC 82

Qy   93 EAPPAPNLLLTCDR---PDLDLRFTIKFQEYSPNLWGHEFRSHHDYYIIATSDGTREGLE 149
```

Db   83 KPHSPDQLRWECSRPFAPHAPEKRFSEKFQRFTPFTLGKEFRQGESYYYIS----------K 133

Qy   150 SL--QGGVCLTRGMKVLLRVGQSPRGGAVPRKPVSEMPERDRGAAHSLEPGKENLPGDP 207

Db   134 PLHHHGQECLRLKVDVV-------GPHGSKNKKKWVEKVEEIEGKMAAGGVHNPSNRLPADD 188

Qy   208 TSNATSRGAEGPLPPPSMPAVAGAAG 233

Db   189 ---------------PIAMIPVVQRSVG 201

RESULT 16
Q9ESQ2

ID   Q9ESQ2   PRELIMINARY;   PRT;   1691 AA.
AC   Q9ESQ2;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Type IV collagen alpha 5 chain.
GN   COL4A5.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20536494; PubMed=10955041;
RA   Saito K., Naito I., Seki T., Oohashi T., Kimura E., Momota R.,
RA   Kishiro Y., Sado Y., Yoshioka H., Ninomiya Y.;
RT   "Differential Expression of Mouse a5(IV) and a6(IV) Collagen Genes in
RT   Epithelial Basement Membranes.";
RL   J. Biochem. 128:427-434(2000).
EMBL;   AB041350; BAB13673.1; -.
MGD;   MGI:88456; Col4a5.
InterPro; IPRO000087; Collagen.
InterPro; IPRO01442; ProcollagnC4.
Pfam;   PF01413; C4; 2.
Pfam;   PF01391; Collagen; 24.
ProDom; PD000007; Collagen; 3.
SMART; SM00923; ProcollagnC4; 1.
KW   Collagen.
SQ   SEQUENCE   1691 AA; 161823 MW; 81340DF19220BFA CRC64;

Query Match   6.3%; Score 153.5; DB 11; Length 1691;
Best Local Similarity   24.6%; Pred. No. 0.0022;
Matches   86; Conservative   29; Mismatches 119; Indels 115; Gaps 18;

Qy   52 PQIGDRLDLLLCPRARPDPGPHSSPNYEFYKLYLVGGAQGRRCEAPPAPNLILTCDRPDLDL 111

Db   442 PQDPSR-DEIC-KAGPPGPPGPGDK--------GLQGERGVKGDKGDTCFNC--------- 484

Qy   112 RFTIKFQEYSPNLWGHEFRSHHDYYIIATSDGTREGLESLQG------GVCLTRGMKV 163

Db   485 -----------GTGISGPGQPGLPLPGPPGPGSLGIPGEKGDK-- 517

Qy   164 LLRVGQS-----PRG-GAVPRKP--- 196

Db   518 ----GQAGTTGPKGLPGIIGPGPGAPGFPGSKGDPGDVLTLPGMKGDKELGFPGAPGLPG 573

Qy   197 ---EPGKENLPGDP-----TSNATSRGAEGPLPPPSMPAVAGAAGGLALL--------- 238

Qy   239 --LLGVAGAGGAMCWRRRAKPSES-RHPG---- 291
Db   574 LPGSPGKDGLPGLLPGPKGPGPGPGSPGLLPGPKGPGPGITFKGERGPPGSPGLPG 633

Qy   292 ELGI-----ALRGGGAADPPFCHYEKVSGD---YGHPVYIVQDGPPQSP 333
Db   634 KGIQGVAGNPGQPGLPGPSKGDPGQTITQPKGLPGSPGERDGSVGLPGPDPGLPGQPGLPG 693

Db   694 ILGSKGEPGIPGIGPPGPPGPRKGPPGIPGPPGPGAPGAPGRMGPDGPPGP 742

RESULT 17
Q9IB91

ID   Q9IB91   PRELIMINARY;   PRT;   1447 AA.
AC   Q9IB91;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Type I collagen alpha 1.
GN   COL1A1.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Goto T., Katada T., Kinoshita T., Kubota H.Y.;
RT   "Expression and characterization of Xenopus type I collagen alpha 1
RT   (COL1A1) during embryonic development.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
EMBL;   AB034701; BAA94972.1; -.
InterPro; IPRO000087; Collagen.
InterPro; IPRO00885; Fib collagen_C.
InterPro; IPRO01007; VWF_C.
Pfam;   PF01410; COLFI; 1.
Pfam;   PF01391; Collagen; 18.
Pfam;   PF00093; VWC; 1.
ProDom; PD000007; Collagen; 2.
ProDom; PD020078; Fib collagen_C; 1.
SMART; SM00038; COLFI; 1.
SMART; SM00214; VWC; 1.
PROSITE; PS01208; VWFC; 1.
KW   Collagen.
SQ   SEQUENCE   1447 AA; 137446 MW; A4A6DD2B41583B38B CRC64;

Query Match   6.2%; Score 151.5; DB 13; Length 1447;
Best Local Similarity   27.2%; Pred. No. 0.0026;
Matches   86; Conservative   13; Mismatches 128; Indels 89; Gaps 16;

Qy   67 PPGPHSSPNYEFYKLYLVG-------------------GAQGRRCE-APPAPNLILTCDRPDLDL 111
Db   756 PPGPPGPGDK-------GRAGPAGPAGPGPTGARGAGPAGPGERCGEPAPPGPAGF--AGPPGAD- 805

Qy   112 RFTIKFQEYSPNLWGHEFRSHHDYYIIATSDGTREGLESLQG------GVCLTRGMK--- 162
Db   806 -----GQPGAKGEQGDSG-----AKGDAGPSGLPGPTGAPGPAGPAGPAGALGSPGPKGARG 851

Qy   163 -------VLLRVG-QSPRGGAVPRKPVSEMPERDRG--------AAHSLEPGKEN 202
Db   852 APGPPGATGPGAAGRVGPPGPSGNAGPPGPSGPAGKEGAKGPRGEETGPAGRSGEPGAAG 911

Qy   203 LPGDPTSNATSRGAGAGPLPPPSMPAVAGAAGGLALL--LLGVAGAGGAMCWRRRAKPSES 261
Db   912 PPGPPGEKG-SPGSDGPAGARGIFPGPQGVAGSRGTVGLPGMRGERGFSGLPGPAGEPGKQ 970

Qy   262 RHPGP-----GSFGRGGSLGLGGGGMGPREABPGELGIALRGGGAADPPFCPHYEKVSGD 317
Db   971 GSSGPSGERGPPGPGSGPPGLGGGPPGESGRGEAPGSEGAPGR-DGAVGP-------KGD 1020

Qy   318 YGHPVYIVQDGPPQSP 333
Db   1021 RGEAA-----GPPGAP 1031

RESULT 18
O17036

ID   O17036   PRELIMINARY;   PRT;   325 AA.
AC   O17036;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical 31.6 kDa protein.
GN   T15B7.4.

```
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   None;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Pauley A., Gattung S.;
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   "Direct Submission.";
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF022985; AAB59961.1; -.
DR   EMBL; AF000487; Collagen.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR002486; Col_cuticle_N.
DR   Pfam; PF01391; Collagen; 3.
DR   Pfam; PF01484; Col_cuticle_N; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   325 AA;   31591 MW;   6538D4337C9EC202 CRC64;

Query Match             6.1%;  Score 149;  DB 5;  Length 325;
Best Local Similarity   26.0%;  Pred. No. 0.00074;
Matches   92;  Conservative  25;  Mismatches 109;  Indels 128;  Gaps  23;

Qy    13 VGALILLGVLGLVSGLSL-----EPVYWNSANKRFQAEGG---YVLYPQIGDRLDLLC 62
Db     1 MSASTLVTVASAASGTAIVCVFTVGMIFNDINSFYDEKIGELKEFKGYEQIA-WQAMI 58

Qy    63 PRARPPGPHSSPNYEFYKLYLVG-------GAQGRRCEA-PPAPNLLLTCDRPLDL 111
Db    59 PTTRPSSGSS--------------FLLGRNKRQAQCNCGAQSRGCPAGPPGP 96

Qy   112 RFTIKFQEYSPNLWGHEFRSHHDYYIIATSDGTREGLESLQG---GVCLTRGMKVLLRVCQ 169
Db    97 ----PGQPGAPGEQGHPG-------LAGQPGSGARINPATGRPGFCIT--- 133

Qy   170 SPRGGAVPRKPVSEMPMERDRGAAHSLEPGKENLPGDPTPSNATSRGAAGGPLPPPSMPAVA 229
Db   134 CPAGAPGPAGP-------PGA---PGPKGNNGQPGAPAQS-GGRGP-PGPRGP--A 175

Qy   230 GAAGGLALILLGVAGAGGAMCWRRRRAKPSESRHPG-PGSFGRGG--SLGLGGGGGM--- 283
Db   176 GDAG-----------SPGQPGHPGSPGNPGCRGGQRSRGLPGPSGRGP 212

Qy   284 -GPREAEPGELGIALRGGAADP-PFCPHYEKVSGDYGHPVYIVQDGPPQSPPN 335
Db   213 PGP-AGGPGQPG---HSGGAGSPGPGGP-----PGPSGQPGHSGNDGVPGAPGN 257
```

RESULT 19

```
Q17208
ID   Q17208           PRELIMINARY;      PRT;   569 AA.
AC   Q17208;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Collagen (fragment).
OS   Bombyx mori (Silk moth).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC   Bombycoidea; Bombycidae; Bombyx.
```

```
OX   NCBI_TaxID=7091;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EUROPEAN 200X300; TISSUE=IMAGINAL WING DISC;
RX   MEDLINE=97148965; PubMed=8995790;
RA   Chareyre P.P., Besson M.M., Fouche J.J., Bosquet G.G.;
RT   "Identification of a Bombyx collagenous protein with multiple short
RT   domains of Gly-Xaa-Yaa repeats: cDNA characterization and regulation
RT   of expression.";
RL   Insect Biochem. Mol. Biol. 26:677-685(1996).
DR   EMBL; Z30348; CAA83002.1; -.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR001419; Glutenin.
DR   Pfam; PF01391; Collagen; 5.
DR   PRINTS; PR00210; GLUTENIN.
FT   NON_TER       1
FT   NON_TER       1
SQ   SEQUENCE   569 AA;   56583 MW;   FE371482F0A3B0AF CRC64;

Query Match             6.1%;  Score 149;  DB 5;  Length 569;
Best Local Similarity   29.3%;  Pred. No. 0.0014;
Matches   51;  Conservative  14;  Mismatches  77;  Indels  32;  Gaps   6;

Qy   171 PRGGAVPRKPVSEMPMERDRGAAHSLEPGKENLPGDPTPSNATSRGAAGGPLPPPSMPAVAG 230
Db     1 PGQGGQPIKPGQ----------PGYPGQPGQPGYPGQP-GQPGAPGQPGQPGQPGTPG 50

Qy   231 AAGGLALILLGVAGAGGAMCWRRRRAKPSESRHPG------PGSFGRGGSLGLGGGGGM 283
Db    51 QAG-----QPGYPGQGG----QPIKPAQPGHPGQPGQPGQPGQPGTPGQPGQPGYPGQPGQ 99

Qy   284 GPREAEPGELGIALRG----GGAADPPPFCPHYEKVSGDYGHPVYIVQDGPPQSP 333
Db   100 PGQPGQPGQGQGGYPGQGGGPGQPINPSGQPGYPGQSGQPGYPGQPVQPGQPGQP 153
```

RESULT 20

```
Q19111
ID   Q19111           PRELIMINARY;      PRT;   316 AA.
AC   Q19111;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   F02D10.1 protein.
GN   F02D10.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Swinburne J.;
RL   Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99069613; PubMed=9851916;
RA   none;
RT   "Genome sequence of the nematode C. elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
DR   EMBL; Z67990; CAA91932.1; -.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR002486; Col_cuticle_N.
DR   Pfam; PF01391; Collagen; 2.
DR   Pfam; PF01484; Col_cuticle_N; 1.
SQ   SEQUENCE   316 AA;   29444 MW;   4D76D5BA07923499 CRC64;

Query Match             6.1%;  Score 148.5;  DB 5;  Length 316;
Best Local Similarity   28.5%;  Pred. No. 0.0078;
Matches   49;  Conservative  12;  Mismatches  78;  Indels  33;  Gaps   6;

Qy   168 GQSPRGGAVPRKPVSEMPMERDRGAAHSLEPGKENLPGDP-------------TSNATSRGA 216
Db   144 GASGKGAAAPCEPSTPPCQ-PCPAGPPGPPGPPGDGTPGEPGPPGPGPPGPAGSPAGPSGPGPAGP 202
```

```
Qy  217  EGPLPPPSMPAVAGAAGGLIALLLLGVAGAGGAMCWRRRAKPSESRHPGP-GSFGRGGSL  275
Db  203  PGPAGPPGNDGPGPGQPGPG--QDGASSAGG-------EAGPGPAGPPGIPAGPAGPDGQS  253

Qy  276  GLGGGGMGPR-----------EABPGELGIALRGGGAADPPFCPHYEKVSG  316
Db  254  GSGSAGGPGPKGPPGPPGAGGQPGSDGNPGTPAGPPGGPGNPGGSGKGICPKYCAIDG  305
```

RESULT 21

```
ID   Q61434                              PRELIMINARY;   PRT;   1140 AA.
AC   Q61434;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Collagen (Fragment).
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Abe N., Muraaki Y., Yoshioka H., Inoue H., Ninomiya Y.;
RT   "Identification of a novel collagen chain represented by extensive
RT   interruptions in the triple-helical region.";
RL   Cell. Mol. Biol. Res. 196:576-582(1993).
DR   EMBL; D17546; BAA04483.1; -.
DR   HSSP; P39061; 1KOE.
DR   MGD; MGI:88449; Col15a1.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR003129; TSPN.
DR   Pfam; PF01391; Collagen; 8.
DR   Pfam; PF02210; TSPN; 1.
DR   ProDom; PD000007; Collagen; 1.
FT   NON_TER      1
SQ   SEQUENCE   1140 AA;  115156 MW;  8B0C7E68623B3BDFE CRC64;

Query Match   6.0%;
Best Local Similarity   23.7%;   Pred. No. 0.0049;
Matches   92;   Conservative   22;   Mismatches 107;   Indels 167;   Gaps 21;

Qy    2  GPPHSGPGVRVGALILLGVLG-LVSGLSLEPYYWNSANKRFQAEGGYVTLYPQIGDRLDL   60
Db  164  GPP--GPAGPQ----------GPAGPVVQSPNSQPV-                         187

Qy   61  LCPRAR--PPGPHSSPNYEFYKLVLVGGAQGRRCE-----------                93
Db  188  --PGAQGPGPQGGPGKD-------GTPGRDGEPGDPGEDGRPGDTGPQGPGPGFPGTPGDVG  237

Qy   94  ------------APPAPNLLLTCDRPDLDLRFTIKFQEYSPNLMGHEFRSHHDYYIIAT   140
Db  238  PKGEKGDPGIGPRGPGP-------PGPPGPSSFRQDKLTFIDME                  274

Qy  141  SDGTREGLBSLQGGVCLTRGMKVLLRVGQSPRGGAVPRKP--VSEMPMERDR-GAAHSLE  197
Db  275  GSGPSGDIESLRG--------PRGFPGPGPPGVPGLPGEPGRFGINGSYA            317

Qy  198  PGKENLPGDPTSNATSRGAEGP--LP-PPSMPAVAGAAGGLIALLLLGVAGAGGAMCWRR  254
Db  318  PGPAGLPGVP------GKEGPPGPGPGPGPPGPPGPPGPPGPPGPPGPPGPPGP         360

Qy  255  RAKPSESRHPGP-GSFGRGGSLGLGGGGMGPREAEPGELGIALRGG--GAADPPFCPHY   311
Db  361  ----GDVGIPGPKGSKGDLGPIGWPGKSGLA---GSPGPVGPGPPGPPGPPGPFAAGF    413

Qy  312  EKVSGDYGHPVYIVQ------DGPPQSP                                  333
Db  414  DDMEGS-GIPLWTTARSSDLQGSPPGSP                                  440
```

RESULT 22

```
ID   Q91718                              PRELIMINARY;   PRT;   1491 AA.
AC   Q91718;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Alpha-1 type II collagen.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Su M.W., Suzuki H.R., Bieker J.J., Solursh M., Ramirez F.;
RX   MEDLINE=92011898; PubMed=1918153;
RT   "Expression of two nonallelic type II procollagen genes during Xenopus
RT   laevis embryogenesis is characterized by stage-specific production of
RT   alternatively spliced transcripts.";
RL   J. Cell Biol. 115:565-575(1991).
DR   EMBL; M63596; AAA49679.1; -.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR001007; VWF_C.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 18.
DR   Pfam; PF00003; VWC; 1.
DR   ProDom; PD000007; Collagen; 3.
DR   ProDom; PD020078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS01208; VWFC; 1.
KW   Collagen.
SQ   SEQUENCE   1491 AA;  142495 MW;  43026FF08FB0314 CRC64;

Query Match   6.0%;
Best Local Similarity   31.1%;   Pred. No. 0.0067;
Matches   56;   Conservative   11;   Mismatches 80;   Indels 33;   Gaps 8;

Qy  171  PRGGAVPRKPVSEMPMERDRG--AAHSLEPGKENLPGDPTSNATSRGAEGPLPPPSMPAVA  229
Db  123  PRGFPGPGPQGPSGEQGPRGERGDKGETGAPGPRGRGDGEP----GTPGNPGPAGPGPGPP  178

Qy  230  GAAGGLIALLLLG--VAGAGGAMCWRRR-------RAKPSESRHPGPGSF-GRGGSLGLG   278
Db  179  GLGGGNFAAQMTGGFDEKAGRAGAQMGVWQGPMGPMGPRGPGPGPGGPGAPGPGPGGNPGEPGES  238

Qy  279  GGGG-MGPR------EAEPGELGIALRGG-----------GAADPPFCPHYEKVSGDYGHP  321
Db  239  GAGGPMGPRGPGPGPSGKPGDDGEAGKPGKSCGERGPPGPGPQGARGPFPGTTPGLPGVKGHRGYP  298
```

RESULT 23

```
ID   Q62001                              PRELIMINARY;   PRT;   1774 AA.
AC   Q62001; Q60672;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Procollagen, type XVIII, alpha 1 precursor (XVIII) collagen
DE   (procollagen, type XVIII, alpha 1) (Alpha-1 type XVIII collagen).
GN   COL18A1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PRIMARY TAIL CULTURE;
RA   Rehm M., Hintikka E., Pihlajaniemi T.;
RX   MEDLINE=94245707; PubMed=8188673;
RT   "Primary structure of the alpha 1 chain of mouse type XVIII collagen,
```

```
RT  partial structure of the corresponding gene, and comparison of the
RT  alpha 1(XVIII) chain with its homologue, the alpha 1(XV) collagen
RT  chain.";
RL  J. Biol. Chem. 269:13929-13935(1994).
RN  [2]
RP  SEQUENCE OF 1-562 FROM N.A.
RX  MEDLINE=94240112; PubMed=8183994;
RA  Rehn M., Pihlajaniemi T.;
RT  "Alpha 1(XVIII), a collagen chain with frequent interruptions in the
RT  collagenous sequence, a distinct tissue distribution, and homology
RT  with type XV collagen.";
RL  Proc. Natl. Acad. Sci. U.S.A. 91:4234-4238(1994).
RN  [3]
RP  SEQUENCE OF 1-562 FROM N.A.
RX  MEDLINE=95181468; PubMed=7876242;
RA  Rehn M., Pihlajaniemi T.;
RT  "Identification of three N-terminal ends of type XVIII collagen chains
RT  and tissue-specific differences in the expression of the corresponding
RT  transcripts. The longest form contains a novel motif homologous to rat
RT  and Drosophila frizzled proteins.";
RL  J. Biol. Chem. 270:4705-4711(1995).
DR  EMBL; U03715; AAC52903.1; -.
DR  EMBL; U03716; AAC52903.1; JOINED.
DR  EMBL; U03718; AAC52903.1; JOINED.
DR  EMBL; U34607; AAC52903.1; JOINED.
DR  EMBL; U34608; AAC52903.1; JOINED.
DR  EMBL; U34609; AAC52903.1; JOINED.
DR  EMBL; U34610; AAC52903.1; JOINED.
DR  EMBL; U34611; AAC52903.1; JOINED.
DR  EMBL; U34612; AAC52903.1; JOINED.
DR  EMBL; U34613; AAC52903.1; JOINED.
DR  EMBL; U11637; AAC52179.1; -.
DR  HSSP; P39061; 1KOE.
DR  MGD; MGI:88451; Col18a1.
DR  InterPro; IPR000087; Collagen.
DR  InterPro; IPR000024; Fz_domain.
DR  InterPro; IPR001791; Laminin_G.
DR  InterPro; IPR003129; TSPN.
DR  Pfam; IPR01391; Collagen; 8.
DR  Pfam; PF01192; Fz; 1.
DR  Pfam; PF02210; TSPN; 1.
DR  ProDom; PD000007; Collagen; 1.
DR  SMART; SM00063; FRI; 1.
DR  SMART; SM00282; LamG; 1.
DR  SMART; SM00210; TSPN; 1.
DR  PROSITE; PS50038; Fz; 1.
KW  Signal.
SQ  SEQUENCE  1774 AA;  182229 MW;  CF4D9BC9E88BF232 CRC64;
```

```
Query Match            6.0%;
Best Local Similarity  23.7%;
Matches    92;   Conservative   22;   Mismatches 107;   Indels 167;   Gaps  21;

Qy      2  GPPHSGPGGVRVGALILLGVLG-LVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDL  60
           |||   |    :                     |: |:
Db    798  GPP---GPAGPQ----------------------GPAGPVVQSPNSQPV-----------  821

Qy     61  LCPRAR-PPGPHSSPNYEFYKLIVLVGGAQGRRCE--------------------------  93
               ||:||  :|        ||| |
Db    822  --PGAQGPPGPQGPPGKD---------------------GTPGRDGEPGDPGEDGRPGDTPQGFPGTPGDVG  871

Qy     94  ---------------------APPAPNLLLTCDRPDLDLRFTIKFQEYSPNLWGHEPRSHDYYIIAT  140
                                  :       |             :|
Db    872  PKGGEKGDPGIGPRGPPGP----------------PGPGPGPSFRQDKLTFIDME  908

Qy    141  SDGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKP--VSEMPMERDR-GAAHSLE  197
                                         |:|:|                    :||:
Db    909  GSGFSGDIESLRG---------------------PRGFPGPPGPPGPVPGLPGEPGPRFGINGSYA  951

Qy    198  PGKENLPGDPTSNATSRGAEGP-----LP-PSMPAVAGAAGGLALILLGVAGAGAMCWRRR  254
                |:|  |   ||      |    |  |   ::   |||  |  ::|    |
Db    952  PGPAGLPGVP--------GKEGPGPGFGPGPGPGPPGPGKEGP-----PGVAGQKGSV---  994
```

```
Qy    255  RAKPSESRHPGP-GSFGRGGSLGLGGGGMGPREAEPGELGIALRGG--GAADPPFCPHY  311
           ||| |||| ||  |: |:               : |    :        ::|  |  |
Db    995  ----GDVGIPGPKGSKGDLGPIGMPGKSGLA---GSFGPGVGGPPGPPGPPGPGFGPAAGF  1047

Qy    312  EKVSGDYGHPVYIVQ--------DGPPQSP  333
            | |||||||    :|         :|::|
Db   1048  DDMEGGS-GIPLWTTARSDSDGLQGPPGSP  1074
```

```
ID   Q90800      PRELIMINARY;      PRT;    675 AA.
AC   Q90800;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Collagen-alpha-3 type IX precursor.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=CARTILAGE;
RX   MEDLINE=92241276; PubMed=1572350;
RA   Brewton R.G., Ouspenskaia M.V., van der Rest M., Mayne R.;
RT   "Cloning of the chicken alpha 3(IX) collagen chain completes the
RT   primary structure of type IX collagen.";
RL   Eur. J. Biochem. 205:443-449(1992).
RL   EMBL; X64712; CAA45967.1; -.
DR   InterPro; IPR000087; Collagen.
DR   Pfam; PF01391; Collagen; 11.
DR   ProDom; PD000007; Collagen; 3.
KW   Collagen; Signal.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22    675       COLLAGEN-ALPHA-3 TYPE IX.
SQ   SEQUENCE   675 AA;   63069 MW;   9509437A339D72D CRC64;
```

```
Query Match            6.0%;
Best Local Similarity  28.8%;
Matches    64;   Conservative    9;   Mismatches  71;   Indels  78;   Gaps  12;

Qy    166  RVGQSPGGGAVPRKPVSEMPMERDRGAAHSLEPGKENLPGDPTSNATSRGAEGPLPPPSM  225
           |:     |                    :   ||   |:|            |
Db     23  RVG---PQGPPGPRGPPGPSGKDGIDG----EPGPSGLPGPP-----------GPKGAPGK  65

Qy    226  PAVAGAAG--GLALILLGVAGAGAMCWRRRRAKPSESRHPG-----------PGSFGRG-  272
                 | |   |     |||  |  |      |    :||            ||:||:|
Db     66  PGAAGEAGLPGLP----GVDGLTGT------DGPPGPNGPPGDRGALGPAGPGPAGKGL  115

Qy    273  ------GSLGLGGGGGM----GPR------------------RAEPGELGIALRGG------  300
                 |: |:|:|     ||                   ||||||||||||
Db    116  PGPPGPGPGPSGLPGGNGPRGPPGPSGLPGFPGPPGPPGLAGIIPEGGGDLQCPPALCP  175

Qy    301  -GAADPPFCPHYEKVSGDYGHPVYIVQDGP-----PQSPPNI  336
                |||:|    ||  ||||||| | :|        |:|||
Db    176  PGPPGPGPGMPGFKGHTGHKGEPGSIGKEGEKGSPGPGPGGI  217
```

```
ID   Q9QZS0      PRELIMINARY;      PRT;    1669 AA.
AC   Q9QZS0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Alpha 3 collagen IV.
GN   COL4A3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=KIDNEY;
RX   MEDLINE=2005334; PubMed=10534397;
RA   Lu W., Phillips C.L., Killen P.D., Hlaing T., Harrison W.R.,
RA   Elder F.F.B., Miner J.H., Overbeek P.A., Meisler M.H.;
RT   "Insertional mutation of the collagen genes col4a3 and col4a4 in a
RT   mouse model of alport syndrome.";
RL   Genomics 61:113-124(1999).
DR   EMBL; AF169387; AAD50449.1; -.
DR   MGD; MGI:104688; Col4a3.
DR   InterPro; IPR001442; Procollagen.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF01413; C4; 2.
DR   Pfam; PF01391; Collagen; 21.
DR   ProDom; PD000007; Collagen; 6.
DR   ProDom; PD003923; ProcollagnC4; 1.
DR   SMART; SM00111; C4; 2.
DR   PROSITE; PS00030; RRM_RNP_1; UNKNOWN_1.
KW   Collagen.
SQ   SEQUENCE   1669 AA;  161769 MW;  30976E59739A47B2 CRC64;

Query Match             6.0%;  Score 146;  DB 11;  Length 1669;
Best Local Similarity   24.3%;  Pred. No. 0.0083;
Matches   97;  Conservative   34;  Mismatches  134;  Indels  134;  Gaps  22;

Qy      4 PHSGPGGVRVGALLLLGVLGL----------------------------VSGLSLEP--VYWNSANKRFQAEGGY 48
            :  :  :          :                              ::::::       :
Db   1008 PHGMPGS------------MGIWGVPGPGRKGTSGLPGLAGRPGRTIHGPQGDK---GEPGY 1056

Qy     49 VLYPQIGDRLDLLCPRARPPGPHSSPNYEFYKLYLVGAQGRRCE--------- 93
                           ||||
Db   1057 SEGARPG------PPGPKGDPG-------LPGDKGKGERGVPGPGPPQQSGPAGP 1097

Qy     94 --APPAPNLLLTCDRP----DLDLRFTIKFQEYSPNLWGHEFRSHHDYYIIATSDGTREG 147
            :  :                    :
Db   1098 DGAPGSPGSPGHPGKPGPAGDLGL----KGQKGPFGPPG-------STGPPGPPG 1141

Qy    148 LESLLQGGVCLI--------TRGMKVLLRVGQSPRGGAVPRKPVSEMPMERDRGAA 193
Db   1142 LPGLPGPMGWRGDQGRDGIPGPPGGEKGETGLLGAYPGPKGS----KGDRGV----- 1194

Qy    194 HSLEPGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAGGLALLLLGVAGAGGAMGWRR 253
Db   1195 ----PGLSGLPG------RKGVWGDVGPGQGPGPGTAGLPGPPGPGLPGAIIPGPKGDRGLPG 1243

Qy    254 RRAKPSESRHPG-PGSFGRG---------GSLGLGGG-GGMGPREAPPGELGIALRGG 300
Db   1244 LRGNGPGEPGPGPGPIGKIKGDKGPMGPPGPGVKGLPGTVGDMGP-PGPPGAPGTPGLPG 1302

Qy    301 GAADPPFCPHYEKYSGDYGHPVY----IVQDGP--PQSPP 334
Db   1303 VRGDPGF-PGPPGIKGEKGNPGFLGPIGHPGPVGPKGPP 1340

RESULT 26
ID   O17805
AC   O17805;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   F14F7.1 protein.
GN   F14F7.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   McMurray A.A.;
RL   Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.

                                     PRELIMINARY;      PRT;      305 AA.
```

```
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99069613; PubMed=9851916;
RA   none;
RT   "Genome sequence of the nematode C.elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
DR   EMBL; Z81503; CAB04111.1; -.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR002486; Col_cuticle_N.
DR   Pfam; PF01484; Col_cuticle_N; 1.
SQ   SEQUENCE   305 AA;  29497 MW;  161BFA90D024C2E6 CRC64;

Query Match             5.9%;  Score 145.5;  DB 5;  Length 305;
Best Local Similarity   22.5%;  Pred. No. 0.0013;
Matches   83;  Conservative   21;  Mismatches  122;  Indels  143;  Gaps  12;

Qy     11 VRVGALLLLG-------VLGLVSGLSLEPVYWN----------SANK----- 40
            |||       |      ||:|:||||:|:|||          ||||
Db      5 VRLKAYRLLAFSAVAFCVLSVVSVCITLPMVYNYVSGMRVRLVQEVSFCKHSANEVFTEV 64

Qy     41 ------RFQAEGGYVLYPQIGDRLDLLCPRARPPGPHSSPNYEFYKLYLVGG 86
Db     65 NFLRASATNSTRSARAHGYGGY--------------AQPGQGGGG------GG 97

Qy     87 AGGRRCRRAPPAPNLLLITCDRPDLDLRFTIKFQEYSPNLWGHEFRSHHDYYIIATSDGTRE 146
Db     98 GQCSTCCRPGPPGCAGGTPGKPG---------------RPGAPGAWG------- 128

Qy    147 GLESLCGGVCLITRGMKVLLRVGQSPRGGAVPRKPVSEMPMERDRGAAHSLEPGKENLPGD 206
            ||:||  |:|:|||||||||||||||||||||||||||
Db    129 MFGNPGKGGSGPGCHPGVVPTPCKPGCPGG-----RPGPPGPPGR 165

Qy    207 PTSN------ATISRGA---------EGPLPPPSMPAVAGAAGGLALLLLGVAGAGAMGW 251
Db    166 PGSDGQPGPATGGGAAPRPGPGPGKGPRGAPGNSGRAGAGPGQPGNDAHGYGGGVGPGPGP 225

Qy    252 RRRRAKPSESRHPGPGSFGRGGSSLGLGGGGGM-----GPREAPGELGIALRGGGAADPPF 307
Db    226 AGPRGAPGPAGHPGSSGGGRGPGPAGEKGAPGQPGRGP-DGHPGGQPGRPGCGSGGSGNRGV 284

Qy    308 CPHYEKVSG 316
Db    285 CPXYCAIDG 293

RESULT 27
ID   Q94620
AC   Q94620;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Cuticle preprocollagen.
GN   COL-2.
OS   Meloidogyne incognita (southern root-knot nematode).
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
OC   Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.
OX   NCBI_TaxID=6306;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=CUTICLE;
RX   MEDLINE=98324411; PubMed=9662034;
RA   Wang T., Deom C.M., Hussey R.S.;
RT   "Identification of a Meloidogyne incognita cuticle collagen gene and
RT   characterization of the developmental expression of three collagen
RT   genes in parasitic stages.";
RL   Mol. Biochem. Parasitol. 93:131-134(1998).
DR   EMBL; U67729; AAC48838.1; -.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR002486; Col_cuticle_N.
DR   Pfam; PF01391; Collagen; 2.
DR   Pfam; PF01484; Col_cuticle_N; 1.

                                     PRELIMINARY;      PRT;      308 AA.
```

```
KW   Collagen.
FT   CHAIN        80       308
SQ   SEQUENCE   308 AA;  29361 MW;  825AEF0249FDFC78 CRC64;      COLLAGEN.

Query Match             5.9%;    Score 145.5;  DB 5;   Length 308;
Best Local Similarity  28.4%;    Pred. No. 0.0013;
Matches   46;  Conservative 14;  Mismatches 67;  Indels 35;  Gaps 6;

Qy    171 PRGGAVPRKPVSEMPERDRGAAHSLEPGKENLFGDPTSNATSRGAEBGPLPPSMPAVAG 230
Db    155 PCPGGQPGCPQPGGPGGPPPGGPPGQPGGPGGQPGGQPGGQASPGEP----------- 204
Qy    231 AAGGLLALLILGVAGAGGAMCWRRRRAKPSSSRHPG--------PGSFGRGGSLGL-GGGGGMGP 285
Db    205 QPGG-------------------------PGQPGAPA-QSQEAGPAPPGPGDAGAPGQPNPPGAGPGQPGPGPGSP 255
Qy    286 R-------------EAEPGELGIALRGGGAADPPFCPHYEKVSG 316
Db    256 KGPPGPPGPGPGNDGSPGQPGPPGTNGGSGSGEKGICPKYCAIDG 297
```

```
RESULT 28
ID   Q9D0D2
AC   Q9D0D2;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Procollagen, type IX, alpha 1.
GN   COL9A1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=EMBRYO;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
KW   Collagen.
SQ   SEQUENCE   680 AA;  64733 MW;  5C0FB3E9882B4D91 CRC64;

Query Match             5.9%;    Score 145;  DB 11;   Length 680;
Best Local Similarity  24.0%;    Pred. No. 0.0035;
Matches   87;  Conservative 22;  Mismatches 131;  Indels 122;  Gaps 17;

Qy      2 GPPHSGPGGVRVGALLLGLVSGLSLEPVYWNSANKRFQAEGGYVLLYPQIGDRLDLL 61
          |||   ||| :|                    | :|
Db      2 GPP---------------------------------------------------------PQGQTG 911
```

```
RESULT 29
ID   Q9NZQ6
AC   Q9NZQ6;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Collagen type V alpha 3 chain.
GN   COL5A3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLACENTA, AND HEART;
RX   MEDLINE=20187594; PubMed=10722718;
RA   Imamura Y., Scott I.C., Greenspan D.S.;
RT   "The pro-alpha3 (V) collagen chain. Complete primary structure,
RT   expression domains in adult and developing tissues, and comparison to
RT   the structures and expression domains of the other types V and XI
RT   procollagen chains.";
RL   J. Biol. Chem. 275:8749-8759(2000).
DR   EMBL; AF177941; AAF59902.1; -.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR003129; TSPN.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 17.
DR   Pfam; PF02210; TSPN; 1.
DR   ProDom; PD000007; Collagen; 3.
DR   ProDom; PD000078; Fib_collagen; 3.
DR   SMART; SM000038; COLFI; 1.
DR   SMART; SM000282; LamG; 1.
DR   SMART; SM002010; TSPN; 1.
SQ   SEQUENCE   1745 AA;  172051 MW;  5E8FF97135397AC1 CRC64;

Query Match             5.9%;    Score 145;  DB 4;   Length 1745;
Best Local Similarity  24.0%;    Pred. No. 0.01;
Matches   88;  Conservative 24;  Mismatches 103;  Indels 152;  Gaps 20;

Qy      2 GPP-HSGPGGVRVGALLLGVLGLVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDL 60
          |||  ||| |||                              | :
Db    884 GPPGHQGKDG-RGHPGQRGELG----------------------------PQGQTG 911
```

```
Qy    61 LCPRARPPGPHSSPNYEYFKLYLVGAGAGRRCE------------APPAPNLLLTCRPDLDLRF 113
            |||||                                         |:|:
Db   912 ------------PPGPAG------------VLPGQGKTGEVGPLGERGPPGP------ 939

Qy   114 TIKFQEYSPNLWGHEFRSHDYYIIATSDGTRSGLESLQGGVCLTRGMKVLLRVGQSPRG 173
                                                  |||||
Db   940 ------------PGPPG------------EQGLPGLEG------REGAKG 959

Qy   174 GAVPRKPVSEMPMERDRGAAAHSLEPGKENLPGDPTSNATSRGAABGPLPP-------PSMPA 227
              :|||:                                          |:|:
Db   960 ELGPPGPLGKEGPAGLRGP-----PGPRGPGPGDGPGTGL-KGDKGPGPPGPVGANGSPGERG 1013

Qy   228 VAGAAGGLALILLGVAGAGGAMCWRRRAKPSESRHPGP-GSPGRGGSLG-LGGGGGMGP 285
            |:|:                                    |||||
Db  1014 PLGPAGGIG--LPGQSGSGEGPVGPAGKGSRGERGPPGPTGKDGIPGPLGPGPGPGAAGP 1071

Qy   286 --REAEPGELGIALRGG--------GAADPPFCPHYEKVSGDYGHPVYIVQDGPP------ 330
             |:|:                     ||||
Db  1072 SGGEEGDKGDVGAPGHKGSKGDKGDAGPPGQP--GIRGPAGHP-----GPPGGADGAQGR 1122

Qy   331 QSPENIY 337
            :||:
Db  1123 RGPPGLF 1129
```

RESULT 30
```
ID   Q25466      PRELIMINARY;     PRT;   309 AA.
AC   Q25466;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   COL-1.
GN   COL-1.
OS   Meloidogyne incognita (southern root-knot nematode).
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
OC   Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.
     NCBI_TaxID=6306;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97164124; PubMed=9010847;
RA   Ray C., Wang T.Y., Hussey R.S.;
RT   "Identification and characterization of the Meloidogyne incognita coll
RT   cuticle collagen gene.";
RL   Mol. Biochem. Parasitol. 83:121-124(1996).
DR   EMBL; U40766; AAC47437.1; -.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR0024B6; Col_cuticle_N.
DR   Pfam; PF01391; Collagen; 2.
DR   Pfam; PF01484; Col_cuticle_N; 1.
SQ   SEQUENCE 309 AA; 0C6C9D5BA76C0916 CRC64;
```

Query Match   28.4%;
Best Local Similarity 5.9%;   Score 144.5; DB 5; Length 309;
Matches   46; Conservative   14; Pred. No. 0.0016; Mismatches   67; Indels   35; Gaps   6;

```
Qy   171 PRGGAVPRKPVSEMPMERDRGAAAHSLEPGKENLPGDPTSNATSRGAEGPLPPPSMPAVAG 230
                   :|:                                     ||:|:
Db   156 PCPGGQPGQPGPGPPGPGPGPGPGPGGPGQPGPGPGQASPGEP------GPAGPPGSPGAPG 205

Qy   231 AAGGLALILLGVAGAGGAMCWRRRAKPSESRHPG-----PGSPGRGGSLGL-GGGGGMGP 285
            :|:|                                      ||:|
Db   206 QPGG---------PGQPGSPA-QSQEAGPAPPGPGPDAGAPGQPGNPGPAGPQPGPGPGGP 256

Qy   286 R----------EAEPGELGIALRGGGAADPPFCPHYEKVSG 316
                      |:|:                   ||||
Db   257 KGPGPPGPGQPGNDAPGQPGTPGTPGGAGEKGICPKYCAIDG 298
```

RESULT 31
```
ID   Q96QB3      PRELIMINARY;     PRT;   1347 AA.
AC   Q96QB3;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Type V preprocollagen alpha 2 chain (Fragment).
GN   COL5A2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
     NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21451029; PubMed=11562270;
RA   Valkkila M., Melkoniemi M., Kvist L., Kuivaniemi H., Tromp G.,
RA   Ala-Kokko L.;
RT   "Genomic organization of the human COL3A1 and COL5A2 genes: COL5A2 has
RT   evolved differently than the other minor fibrillar collagen genes.";
RL   Matrix Biol. 20:357-366(2001).
KW   Collagen.
FT   NON_TER   1   1
FT   NON_TER   1347   1347
SQ   SEQUENCE 1347 AA; 128822 MW; F25F20E6BB23A17C CRC64;
```

Query Match   33.0%;
Best Local Similarity 5.9%;   Score 144.5; DB 4; Length 1347;
Matches   61; Conservative   12; Pred. No. 0.0085; Mismatches   77; Indels   35; Gaps   11;

```
Qy   169 QSPRG--GAV-PRKPVSEMPMERDRGAAAHSLEPGKENLPG----------DPTSNATSRGAE 217
            |:|:                                                :||:
Db   351 RGPRGPDGPTVGPPGPVGERGAPGPNGRGF------PGSDGLPGPKGPGAQGERGPGPGSPGKGSQ 405

Qy   218 GPLPPPSMPAVAGAAGGLALILLGVAGAGGAMCWRRRAKPSESRHPG-PGSFG-RG--G 273
             |:|:                                           :||:
Db   406 GDPGPGGPCEPGLPGARG-----LTGNPGVGQPEGKLGPLGAPGEDGRPGPPGGSIGIRGQPG 460

Qy   274 SLGLGGGGGMGPREAEPGELGIA----LRGGGAADPPFCPHYEKVSGDYGHPVYIVQDGP 329
              |:|:                     ||||
Db   461 SMGLPGPKGSGSGDPGKPGEAGNAGVPGCQRGAPGKDGEVGP------SGPVGPPGLAGERG 514

Qy   330 PQSPP 334
            :|||
Db   515 EQGPP 519
```

RESULT 32
```
ID   Q90W37      PRELIMINARY;     PRT;   1420 AA.
AC   Q90W37;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Alpha 1 type IIA collagen precursor.
GN   COL2A1.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
     NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=STERNAL;
RA   Caixia X., Yongzhi X., Siqi G., Yiying S.;
```

```
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY046949; AAK98621.1; -.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 18.
DR   ProDom; PD000007; Collagen; 6.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   Collagen; Signal.
KW   Collagen; Signal.
SQ   SEQUENCE   1420 AA;  134999 MW;  88D9AA817F214FF5 CRC64;
FT   SIGNAL        1     25       POTENTIAL.
     SIGNAL
     SEQUENCE
```

Query Match   5.9%; Score 144.5; DB 13; Length 1420;
Best Local Similarity 28.1%; Pred. No. 0.009;
Matches   65; Conservative   11; Mismatches   90; Indels   65; Gaps   10;

```
Qy   147  GLESLQG--GVCLTRGMKVLLRVGQSPRGGAVPRKP------------VSEMMER------  188
Db   644  GAQGLQGPRGLPGTPGT----DGPKGATGPAGPNGAQGPPGLQGMPGEERGAAGIAGL  696

Qy   189  --DRGAAHSLEPGKKNLPGDPTSNATSRGAEGPLPPPSMPAVAGAAGGLALLLLLGVAGAG  246
Db   697  KGDRGDVG--EKGPEGAPG----KDGARGLTGPIGPPGPPGPAGPNGEKGESG--PPGPSGAA  748

Qy   247  GAMCWRRRAKPSESRHPG-PGSFGRGGSLGLGGGGMGPREAEPGELGIALRGG-----  300
Db   749  GA-------RGAPGERGEPGAPGPAGPFAGPGPADGGPGAGKEGBPGGVKGKDAGAPGPQGPS  802

Qy   301  -----------------GAADPPFCPHYEKVSGDYGHPVYIVQDGPPQSP  333
Db   803  GAPGPQGPTGVTGPPGKGARGAQGP2GATGPFPGAAGBVGPPGPNGNPGPGPP  853
```

```
RESULT 33
ID   Q9F342            PRELIMINARY;     PRT;   775 AA.
AC   Q9F342;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein SCO5128.
GN   SCO5128 OR SC9B12.13.
OS   Streptomyces coelicolor.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2) / M145;
RA   Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA   Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA   Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA   Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA   Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA   Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA   Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA   Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA   Hopwood D.A.;
RT   "Complete genome sequence of the model actinomycete Streptomyces
RT   coelicolor A3(2).";
RL   Nature 417:141-147(2002).
DR   EMBL; AL391751; CAC05758.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   775 AA;  75820 MW;  E5032698A20A0711 CRC64;
```

Query Match   5.9%; Score 143.5; DB 16; Length 775;
Best Local Similarity 23.5%; Pred. No. 0.0054;
Matches   94; Conservative   23; Mismatches   130; Indels   153; Gaps   19;

```
Qy    3  PPHGGPG------------------GVRVGALLLLGVLGLVSGLSLEPVYWNSANKRFQAEGGYVL  50
Db   133  PRKSGPGCKGTGQGGSSTNGAGLPAGSAPTAGTPGPAAGARGAGPAGPAGASGAG----------  181
```

```
RESULT 34
ID   Q26640            PRELIMINARY;     PRT;   1747 AA.
AC   Q26640;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Alpha2(IV)-like collagen.
DE   COL04ALPHA.
OS   Strongylocentrotus purpuratus (purple sea urchin).
OC   Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC   Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
OC   Strongylocentrotus.
OX   NCBI_TaxID=7668;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94230414; PubMed=8175744;
RA   Ramirez F.;
RA   Exposito J.Y., Suzuki H., Geurjon C., Garrone R., Solursh M.,
RT   "Identification of a cell lineage-specific gene coding for a sea
RT   urchin .alpha.2(IV)-like collagen chain.";
RL   J. Biol. Chem. 269:13167-13171(1994).
DR   EMBL; X76730; CAA54146.1; -.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR001442; Procollagen.
DR   Pfam; PF01413; C4; 2.
DR   Pfam; PF01391; Collagen; 21.
DR   ProDom; PD000007; Collagen; 6.
DR   ProDom; PD003923; ProcollagnC4; 1.
DR   SMART; SM00111; C4; 2.
KW   Collagen.
SQ   SEQUENCE   1747 AA;  173312 MW;  EE722E8783948996 CRC64;
```

Query Match   5.8%; Score 143; DB 5; Length 1747;
Best Local Similarity 24.1%; Pred. No. 0.015;
Matches   84; Conservative   24; Mismatches   159; Indels   82; Gaps   14;

```
Qy    7  GPGGVRVGALLLLGVLGLVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDLLCPRAR  66
Db   266  GRPGMDG---MKGATGEVGDLGSYGDYGEKLPGYSGERGPPGNPGITPGLLGMNGEKGR  321

Qy   67  PPGPHSSPNYEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDLDLRFTIKFQEYSPNLWG  126
Db  322  -DGPGQGRPGDVGYK--GPSGDYGMGDGPAPDI----------------  352

Qy  127  HEFRSHHDYYIATSDGTREGLESLQGGVCLTRGMKVLLRVGQ----SPRGGAVPRKPVS  182
Db  353  ------EEIVGVGPGPDGEGPPGNPGPGFSGAIGLFGDGDQGTGPGTPGPMGGPGPGPTG  405
```

Qy   51 YPQIGDRLDLLCPRARPPGPHSSPNYEFYKLYLVGGAQ--------GRRCEA---PPAPNLL 101
Db 182 ------------ARPGGGGSGRPGGVGSMNAPGGARSGSAGGGDATGGPVAP--- 224

Qy 102 LTCDRPDLDLRFIKFQEYSPNLWGHEFRSHHDYYIIATSDGTREGLESLQGGVCLTRGM 161
Db 225 ------GH------------GGGTGSFDVTEALAAGPLGGNGN 249

Qy 162 KVLLRVGQSPR----GGAVPRKPEVSEMPMERDRGAAHSLEPGKKENLPGDPTSNATSRGAE 217
Db 250 GNGNGNGSGPRPGTDAGSEPRR--DDLPYFAGEGRADQAPGGQON--CLNGQSPYGTGPQ 305

Qy 218 GPLPPPSMPAV------------AGAAGGLALLLLGVAGAGGAMCWRRRAKPSESRHP---- 264
Db 306 GPAGPTSGPATGDSRIITPPGAGDLGG------LGTAGGPAA--------PGTQREPAPNG 351

Qy 265 --GPGSF-GRGGSLGLGLGGGGMGPREAEPGELGIALRGGGAAD----------PPFCPH 310
Db 352 LRGPEGSLGGPGPGGGPGGPGGPGRPAGSGGPDAARLGP----GGGLSDDTAILTPQRPVPPGAAN 405

Qy 311 YEKVSGD---YGHPVYIVQ-----------DGP-PQSPNI 336
Db 406 PDNISGNTVTSGIPVVPGGERTAPFPAGSGDGRPHTPKL 445

```
Qy    183  EMPMERDRGAAHSLEPGKENLPGDP-------TSNATSRGAEGPL-------PPPSMPAVAGAA 232
                :.:   :                                           .:   :
Db    406  SQCGEGRDGEGKGEGCRPGIFGPFQGETGERGNDGNFGERGEBKGNMGESRGRGPPGCMDGSRGRQ 465

Qy    233  GGLALLLGVAGAGGAMCWRRRAKPSESRHPG--------PGSFGRGGSLGLGGGGGMGP 285
              :.:  ..::::  .    .:                 :.::   : ::  :   :
Db    466  G------PWGQKGPGRGP----PGRAGPAGPARSGNQSFCGFPGDTGNTGLKGMRGIGGALG 516

Qy    286  REAPPGELGIALRGGGAADPPFCPHYEK--VSGGYGHPVYIVQDGPPQSP 333
           ::  :.:    :  ::     ::  :    ..:  : ::  .:  ::  :
Db    517  RDGRPGSKG---ELGGIC--PFCGPGIKGYPGDRGYP------GDPGSP 554
```

```
RESULT 35
ID    O76368             PRELIMINARY;       PRT;      614 AA.
AC    O76368;
DT    01-NOV-1998 (TrEMBLrel. 08, Created)
DT    01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT    01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE    Hypothetical 61.3 kDa protein.
GN    F29C4.8.
OS    Caenorhabditis elegans.
OC    Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC    Rhabditidae; Peloderinae; Caenorhabditis.
OX    NCBI_TaxID=6239;
RN    [1]
RP    SEQUENCE FROM N.A.
RC    STRAIN=BRISTOL N2;
RX    MEDLINE=99069613; PubMed=9851916;
RA    Bradshaw H., Clarke K.;
RT    "The sequence of C. elegans cosmid F29C4.";
RL    Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
RN    [2]
RP    SEQUENCE FROM N.A.
RC    STRAIN=BRISTOL N2;
RA    Waterston R.;
RT    "Genome sequence of the nematode C. elegans: a platform for
RT    investigating biology. The C. elegans Sequencing Consortium.";
RL    Science 282:2012-2018(1998).
RN    [3]
RP    SEQUENCE FROM N.A.
RC    STRAIN=BRISTOL N2;
RA    Waterston R.;
RT    "Direct Submission.";
RL    Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR    EMBL; AF067616; AAC19194.1; -.
DR    InterPro; IPR000087; Collagen.
DR    Pfam; PF01391; Collagen; 8.
DR    ProDom; PD000007; Collagen; 2.
KW    Collagen; Hypothetical protein.
SQ    SEQUENCE   614 AA;  61298 MW;  A76B637B29CC58DA CRC64;

Query Match             5.8%;  Score 142;  DB 5;  Length 614;
Best Local Similarity   27.2%;
Matches    87;  Conservative   19;  Mismatches 110;  Indels 104;  Gaps 21;

Qy     57  RLDLLCPRARPP----GPHSSPNYEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDLDRF 113
              :.:                :..:.    . : :  :.  .      ::  :
Db     24  RNSCMCP-AGPPGRGPVGPP----------GLPGL---PAPYY---RRPRVPL-- 60

Qy    114  TIKPQEYSPNLWGHEFRSHHDYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQSPRG 173
           .:  .   :.. :  .     : .      . .:.   :        ..   :.:
Db     61  ------SNNLDESISRKMRAFGMLYSPDG----QALQ------LRGM------PGPPG 96

Qy    174  GAVPR---------KPVSEMPMERDRGAAHSLEPGKENLPGDPTSNATSRGAEGPLPPPSMPAVA 229
              ..:         : :                      :  : .   :.     ::  :
Db     97  PAGPKGLRGYPGPPGPIGLD-------GPRGLPGTPGSKG-DRGERGPLGPPGPPGPK 146

Qy    230  GAAGGLALLLGVAGAGGAMCWRRRAK--PSESRHPG-PGSFGRGGSLGLGGGGGM-- 283
           : .: . :.:  ..: :       .   :  :.   :::  :    ::
Db    147  GDRGVMTGPYVGPHAGPGPMSHHTNMGNVLPGPGPGPPGPGPAGRDGRHGLKGDRGLPGF 206
```

```
Qy    284  ------GPREAPPGELGIALRGG--GAADPP-------------FCPHYEKVSGGDY-----GH 320
                                                           :    :         :
Db    207  DGESKIGPK-----GEYGSPGRDGIPGARGPPGERGEKGDTAFLSTYPRVASSSTASSPGP 262

Qy    321  PVYIVQDGPPQSPFPNIYTS 340
                : ::. :  :
Db    263  P--------GPP-GPPGVCHAS 275
```

```
RESULT 36
ID    Q96D07             PRELIMINARY;       PRT;      744 AA.
AC    Q96D07;
DT    01-DEC-2001 (TrEMBLrel. 19, Created)
DT    01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT    01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE    Hypothetical 73.4 kDa protein.
OS    Homo sapiens (Human).
OC    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX    NCBI_TaxID=9606;
RN    [1]
RP    SEQUENCE FROM N.A.
RC    TISSUE=LUNG;
RA    Strausberg R.;
RL    Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
DR    EMBL; BC013581; AAH13581.1; -.
DR    InterPro; IPR001073; C1q.
DR    InterPro; IPR000087; Collagen.
DR    Pfam; PF00386; C1q; 1.
DR    Pfam; PF01391; Collagen; 7.
DR    ProDom; PD000007; Collagen; 2.
DR    PROSITE; PS01109; C1Q; UNKNOWN 1.
DR    PROSITE; PS01113; C1Q; UNKNOWN 1.
KW    Collagen; Hypothetical protein.
SQ    SEQUENCE   744 AA;  73364 MW;  2BC1B0955DE2C9A3 CRC64;

Query Match             5.8%;  Score 142;  DB 4;  Length 744;
Best Local Similarity   23.3%;
Matches    72;  Conservative   16;  Mismatches 99;  Indels 122;  Gaps 13;

Qy     67  PPGPHSSPNYEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDLDLRFTIKFQEYSPNLWG 126
              ::.::                :. :.:  .       ::   :  : :
Db    198  PPGPHGLPG-------IGKPGPGPLPGGPGPK---------------- 222

Qy    127  HEFRSHHDYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKPVSEMPM 186
                                    .:  .   :            .  .  ::
Db    223  ------------GRGPKGLPGPCG-----LRGPKGDGFGPMGPGAPGVRGP- 256

Qy    187  ERDRGAAHSLEPGKENLPGDPTSNATS-RGAEGPLPPPSMPAVAGAAGGLALLLL---G 241
           ::  :.   :  : .:     .   :    :  :   :.   :  .  : . :       :
Db    257  PGMHG-PPGPVGLPGVGKPGVTGFPGPGPGDGPLPGKPGAPGEPGPGPGICGVPGVQGPPG 311

Qy    242  VAGAGGAMCWRRRAKPSESRHPG--------------------------PGSFGRGGS 274
           : :::.:                                           ::  :.:
Db    312  IPGIG-------KPGQDGIPGQGPGGKGEQGLPGLPGPPGPLPGIGKPGFPGPKGD 361

Qy    275  LGLGG-GGGMGPREAEPGELGIALRGGGGAADPPFCPHYEKVSGDYGHPVYIVQDGP---- 329
           : :: . ::   :.::    :    :            :    .  : ::  .. .
Db    362  RGMGGVPGALGPR-GEKGPIGAPGIGG----PPGEPGLPGIPGFMGPPGAIGFPGPKGBG 416

Qy    330  ----PQSPP 334
               ::.::
Db    417  GIVGPGQPP 425
```

```
RESULT 37
ID    Q9IAU4             PRELIMINARY;       PRT;      1835 AA.
AC    Q9IAU4;
DT    01-OCT-2000 (TrEMBLrel. 15, Created)
DT    01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT    01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
```

```
DE   Alpha 1 (V) collagen.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BREED WHITE LEGHORN;
RX   MEDLINE=20068042; PubMed=10601735;
RA   Gordon M.K., Marchant J.K., Foley J.W., Igoe F., Gibney E.P.,
RA   Nah H.D., Barembaum M., Myers J.C., Rodriguez E., Dublet B.,
RA   van der Rest M., Linsenmayer T.F., Upholt W.B., Birk D.E.;
RT   "Complete primary structure of the chicken alpha1(V) collagen chain.";
RL   Matrix Biol. 18:481-486(1999).
DR   EMBL; AF137273; AAF28099.1; -.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001791; LamInin_G.
DR   InterPro; IPR001230; Prenyl_site.
DR   InterPro; IPR003129; TSPN.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 19.
DR   Pfam; PF02210; TSPN; 1.
DR   ProDom; PD000007; Collagen; 2.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00210; TSPN; 1.
DR   SMART; SM00282; LamG; 1.
PROSITE; PS00094; PRENYLATION_UNKNOWN_1.
KW   Collagen.
SQ   SEQUENCE   1835 AA;  184234 MW;  D05B9D71022D44B2 CRC64;

Query Match             5.8%;  Score 142;  DB 13;  Length 1835;
Best Local Similarity  25.7%;  Pred. No. 0.019;
Matches   78;  Conservative  15;  Mismatches  82;  Indels 128;  Gaps  18;

Qy     68  PGPHSSPNYEFYKLYLVGGAQG-----------RRCEAPPAPNLLTCDRPDLDLRFTIKF 117
                                           :||   |||| |  ||
Db    566  PGPMGPPG-------------SCGLKGEVGEMGPQGPRGIQGPPGP-----AGKP----- 602

Qy    118  QEYSPNLWGHEFRSHHDYYIIATSDGTR------------EGLESLQGGVCLTRGMKV 163
           | ||
Db    603  -----------AGSDGARGMPGQTGPKGDRGPDGLAGLPG----------- 636

Qy    164  LLRVGQOSPRGGAVPRKPVSEMPMERDRGAAHSLEPGKENLPGDPTSNATSRGAAEGLPPP 223
                                    | ||| |
Db    637  -----EKGNRGEPGPHGPPGPGPPGEDGEBRG----DDGEVGPRGPLPGPP----GPRGLLGPKGPP 686

Qy    224  SMPAVAGAAGGLALILLGVAGAGAGAMCWRRRAKPSESSRHPG-PGSFGRGGSLGL-GGGG 281
Db    687  GPPGPPGVAG-----MDGQTGPKGNV-----------GPQGEPGPGQGGNPGAQGLPGPQG 732

Qy    282  GMGPREAEPGELGIALRGG-----GAADPPFCPHYEKVSGDYGHPVVIVQDGPP-----Q 331
Db    733  PIGP-----PGEKGPLGKPGLPGMPGADGPP----------GHP--GKEGPPGEKGSQ 773

Qy    332  SPP 334
Db    774  GPP 776
```

```
OC   Trichostrongyloidea; Haemonchidae; Ostertagiinae; Teladorsagia.
OX   NCBI_TaxID=45464;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97039674; PubMed=8885226;
RA   Johnstone I.L., Shafi Y., Mjaeed A., Barry J.D.;
RT   "Cuticular collagen genes from the parasitic nematode Ostertagia
RT   circumcincta.";
RL   Mol. Biochem. Parasitol. 80:103-112(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Johnstone I.L.;
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
SEQUENCE FROM N.A.
RX   MEDLINE=20017523; PubMed=10551361;
RA   Britton C., Redmond D.L., Knox D.P., McKerrow J.H., Barry J.D.;
RT   "Identification of promoter elements of parasite nematode genes in
RT   transgenic Caenorhabditis elegans.";
RL   Mol. Biochem. Parasitol. 103:171-181(1999).
DR   EMBL; X96731; CAA65506.1; -.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR002486; Col_cuticle_N.
DR   Pfam; PF01391; Collagen; 3.
DR   Pfam; PF01484; Col_cuticle_N; 1.
KW   Collagen.
SQ   SEQUENCE   284 AA;  27534 MW;  60775B7D2CC40C77 CRC64;

Query Match             5.8%;  Score 141;  DB 5;  Length 284;
Best Local Similarity  27.2%;  Pred. No. 0.0027;
Matches   61;  Conservative  15;  Mismatches  86;  Indels  62;  Gaps  12;

Qy    121  SPNLWGHEFRSHHDYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKP 180
Db     62  SGNIWREVTRTQ-----VLAAP---TRKARQS-GGGCC----------GCGVSPAGPGGP--- 102

Qy    181  VSEMPMERDRGAAHSLEPGKENLPGDPTSNATSRGAEGLP--------PPSMPAVAGAA 232
Db    103  -------------PGQDGPGDGPGRPGPGDGPPATPAPDVRPCFNCPPGPPGPGAGAP 151

Qy    233  GGLALILLGVAGAGAGAMCWRRRAKPSESSRHP-GPGSFGRGGSLGLGGGGMGPREABPG 291
Db    152  GG-----RGQPGSPGS-----DGQPGNAGNPGGPGPGPIGPPGPGNAGGPG---NAGSPG 197

Qy    292  ELGIALRGGGGAADPPFCPHYEKVSGDYGHPVVIVQDGPPPQSPPN 335
Db    198  APGTLTQGPGSGQGPGPP-------GPQGPP----GPDGQPGPGN 232
```

```
RESULT 38
ID   Q25582      PRELIMINARY;      PRT;      284 AA.
AC   Q25582;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Cuticular collagen.
GN   COLOST-1.
OS   Teladorsagia circumcincta.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;
```

```
RESULT 39
ID   Q25581      PRELIMINARY;      PRT;      284 AA.
AC   Q25581;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Cuticular collagen.
GN   COLOST-2.
OS   Teladorsagia circumcincta.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;
OC   Trichostrongyloidea; Haemonchidae; Ostertagiinae; Teladorsagia;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97039674; PubMed=8885226;
RA   Johnstone I.L., Shafi Y., Mjaeed A., Barry J.D.;
RT   "Cuticular collagen genes from the parasitic nematode Ostertagia
RT   circumcincta.";
RL   Mol. Biochem. Parasitol. 80:103-112(1996).
DR   EMBL; X96732; CAA65507.1; -.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR002486; Col_cuticle_N.
DR   Pfam; PF01391; Collagen;
```

```
DR   Pfam; PF01484; Col_cuticle_N; 1.
KW   Collagen.
SQ   SEQUENCE   284 AA;  27560 MW;  7B76E86D2D41D76 CRC64;

Query Match       5.8%;  Score 141;  DB 5;  Length 284;
Best Local Similarity   27.2%;  Pred. No. 0.0027;
Matches   61;  Conservative  15;  Mismatches  86;  Indels  62;  Gaps  12;

Qy  121 SPNLWGHEFRSHHDYYIIATSDGTREGLESLQGGVCLTRGMKYLLRVGQSPRGGAVPRKP 180
          :||.||  ..   |||  :   :|  |.||.|.||.:: |.|  ||.: :
Db   62 SGNIWREVTRTQ----VLLAAP--TRKARQS-GGGCC--------------GCGVSPAGPGP--- 102

Qy  181 VSEMPMERDRGAAHSLEPGKENLPGDPTSNATSRGAEGPLP----------PPSMPAVAGA 232
          |:|  |  ..     |||| |  || |.|:  |||  |||     ...  | ||| ||||
Db  103 ----------PGQDGPQPGSDGRPGDGPQPPATPAPDVRPCFNCPGPGPPAGAP------ 151

Qy  233 GGLALLLLGVAGAGGAMCWRRRAKPSESRHF-GPGSFGRGGSLGLGGGGMGPRREAEPG 291
            :| |:   |||  |  || | |       | :|   |||| |.|  |.|.| :|:
Db  152 GG-------RGQPGSPGS------DGQPGNAGNPGGPGPGPGPGNAGQPG---NPGSPG 197

Qy  292 ELGIALRGGGAADPPFCPHYEKVSGDYDYGHPVVIVQDGPPQSPPN 335
           :|   |||   |  |   |  ...   |   |  :|  |||| |
Db  198 APGTLTQGPGSQPPGPP------GPQGPP----GPDGQPGPGN 232
```

```
RESULT 40
O17038
ID   O17038     PRELIMINARY;  PRT;  319 AA.
AC   O17038;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical 31.3 kDa protein.
GN   T18B7.3.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99069613; PubMed=9851916;
RA   None.
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Pauley A., Gattung S.;
RT   "The sequence of C. elegans cosmid T18B7.";
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Waterston R.;
RT   "Direct Submission.";
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF022985; AAB69959.1; -.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR002486; Col_cuticle_N.
DR   Pfam; PF01391; Collagen; 3.
DR   Pfam; PF01484; Col_cuticle_N; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   319 AA;  31292 MW;  5242AEC668AB488 CRC64;

Query Match       5.8%;  Score 141;  DB 5;  Length 319;
Best Local Similarity   25.6%;  Pred. No. 0.0031;
Matches   90;  Conservative  27;  Mismatches 111;  Indels 124;  Gaps  23;

Qy   13 VGALLLLGVLGLVSGLSL-----EPVVWNSANKRFQAEGG--YYLYPIQGDRLLLC 62
          :|| |:::.||.:   |:         :        |:  || .:|:||||||||
Db    1 MSASTLIVTVASAASGIAIVVCVPTVGMIFNDINSPYDEKIGELKEFKGYEQIA--NQAMI 58
```

```
Qy   63 PRARPGPHSSPNYEFYKLYLVG----------GAQGRRCEA-PPAPNLLLTCDRPDLDL 111
        |        .|:|             ...                           :   :
Db   59 PTTRPSSGSS-------------FLLGRNKRQAECNCGEQOSRGCPAGPGPGPP--GQPG--- 101

Qy  112 RFTIKPQEYGPNLWGHEFRSHHDYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQSP 171
          :|  |   .|:                                       |:
Db  102 ---ARGEAGLPGIAGO------------PGSGARINPATGRPGFCIT------------- 135

Qy  172 RGGAVPRKPVSEMPMERDRGAAHSLEPGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGA 231
        .||  | |   | || |           ||  | |  ...     ||  | |||| |||
Db  136 AGAPGPAGP-----------PGA-----PGPGKNNGQPGAPAQS-GGRGP--PGPRGP--AGD 177

Qy  232 AGGLALLLLGVAGAGGAMCWRRRAKPSESRHPG-PGSFGRGG--SLGLGGGGG 284
        ||   :|     | |.|                | :|   |||   |.|.||||
Db  178 AG-------------------SPGQPGHPGSPGNPGRGGGRNGTPGASGRPGPGQG 214

Qy  285 PREAEPGELGIALRGGGAADP-PFCPHYEKVSGDYDYGHPVYIVQGPPQSPPN 335
        |        :|   |||   | |   |  ...       |  :|  |||| |
Db  215 PAGA-PGQPG---RSGGAGTPGPQGP-------PGPSGQPGHSGNDGVPGTPGN 257
```

```
RESULT 41
O93485
ID   O93485     PRELIMINARY;  PRT;  809 AA.
AC   O93485;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Alpha 1 type I collagen (Fragment).
OS   Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC   Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
OX   NCBI_TaxID=8022;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Fibroblast;
RA   Saito M., Kuniaski N., Hirono I., Aoki T., Ishida M., Urano N.,
RA   Kimura S.;
RT   "Partial characterization of cDNA clones encoding the three distinct
RT   pro alpha chains of type I collagen from rainbow trout.";
RL   Fisheries Sci. 64:780-786(1998).
DR   EMBL; AB008373; BAA33380.1; -.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 9.
DR   ProDom; PD000007; Collagen; 1.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
KW   Collagen.
FT   NON_TER       1      1
SQ   SEQUENCE   809 AA;  78164 MW;  68C056A7640FCA81 CRC64;

Query Match       5.8%;  Score 141;  DB 13;  Length 809;
Best Local Similarity   27.4%;  Pred. No. 0.0089;
Matches   61;  Conservative  15;  Mismatches  93;  Indels  54;  Gaps   9;

Qy  139 ATSDGTREGLESLQG-----GVCLTRGMKVLLRVGQSPRGGAVPRKPVSEMPMERDRGAA 193
        | | |   |  |  .     || |  | | |     |  |   | |   | ||.||
Db   98 ADGAGGKDGVRGMTGPIGPNGPAGSPGDK-----GETGAPGAVGPSGARGAPGERGEBSGA 152

Qy  194 HSLEPGKENLPGDPTSNAT-----------SRGAEGPLPPPSMPAVAGAAGGLALLL 239
          |||| |  | |                  ||| :||| |||| ||| |||| |:||
Db  153 ----PGPAGFRAGPPGGDGPGAKGEAGDNGAKGDGCGAQGPAGPTGAGPGQGPAGN---- 203

Qy  240 LGVAGAGGAMCWRRRAKPSESRHPG------PGSFGRGGSLGLGGGGMGPRREAEPGE 292
         |  | | |                      |   |  |:   || | | |
Db  204 TGAKGARGA----AGPGGATGFPGAAGRFGPGPGPGPGSGNNGPPGGTPGGKEGQKGNRGE 257

Qy  293 LGIALRGG--GAADPPFCPHYEKVSGDYGHPVYIVQDGPPQSP 333
         |       .||                              |
Db  293 ...
```

Db   258 TGPAGRPGELGAAGPP--------GPKGEKGQPGGDGPNGPSGTP 294

RESULT 42
ID   Q91C0   PRELIMINARY;   PRT;   1449 AA.
AC   Q91C0;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Collagen al(I).
GN   CO11A1.
OS   Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC   Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
OX   NCBI_TaxID=8022;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21257802; PubMed=11358497;
RA   Saito M., Takenouchi Y., Kunisaki N., Kimura S.;
RT   "Complete primary structure of rainbow trout type I collagen
RT   consisting of al(I)a2(I)a3(I) heterotrimers.";
RL   Eur. J. Biochem. 268:2817-2827(2001).
DR   EMBL; AB052835; BAB55661.1; -.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001007; VWF_C.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 18.
DR   Pfam; PF00093; vwc; 1.
DR   ProDom; PD000007; Collagen; 2.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   PROSITE; PS01208; VWFC; UNKNOWN_1.
SQ   SEQUENCE 1449 AA; 137117 MW; 62EBF8A7BFD652B8 CRC64;

Query Match   5.8%; Score 141; DB 13; Length 1449;
Best Local Similarity 27.4%; Pred. No. 0.017;
Matches   61; Conservative 15; Mismatches 93; Indels 54; Gaps 9;

Qy 139 ATSDGTREGLESLQG------GVCLTRGMKVLLRVGQSPRGGAVPRKEVSEMPMERDRGAA 193
Db 738 ADGAGGKDCGVRGMTGPIGPNGPAGSPGDK-----GETGAPGAVGPSGARGAPGERGESGA 792

Qy 194 HSLEPGKENLPGDPTSNAT----------------SRGAEGPLPPSMPAVAGAAGGLALLL 239
Db 793 -----PGPAGFAGPPGPGGDGQOPGAKGEAGDNGAKGDGGAQGPAGPTGAPGPQGPPAGN----- 843

Qy 240 LGVAGAGGAMCWRRRARKSESEHPG--------PGSFGRGGSLGLGGGGMGPREAEPGE 292
Db 844 TGAKGARGA-------AGPPGATGPPGAAGRFGPPGPGGPSGNNGPPGPTPGPGGKEGQKGNRGE 897

Qy 293 LGIALRGG--GAADPPPCPHYEKVSGDYGHPVYIVQDGPPQSP 333
Db 898 TGPAGRPGELGAAGPP--------GPKGEKGQPGDGPNGPSGTP 934

RESULT 43
ID   Q90YC5   PRELIMINARY;   PRT;   219 AA.
AC   Q90YC5;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   EPHRIN-A3.
GN   EPHRIN-A3.
OS   Brachydanio rerio (Zebrafish) (Zebra danio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21412237; PubMed=11520665;
RA   Hirate Y., Mieda M., Harada T., Yamasu K., Okamoto H.;
RT   "Identification of ephrin-A3 and novel genes specific to the midbrain-
RT   MHB in embryonic zebrafish by ordered differential display.";
RL   Mech. Dev. 107:83-96(2001).
DR   EMBL; AB051678; BAB55891.1; -.
DR   InterPro; IPR001799; Ephrin.
DR   Pfam; PF00812; Ephrin; 1.
DR   ProDom; PD002533; Ephrin; 1.
DR   PROSITE; PS01299; EPHRIN; 1.
SQ   SEQUENCE 219 AA; 25146 MW; 7191927E03F8EA01 CRC64;

Query Match   5.7%; Score 140.5; DB 13; Length 219;
Best Local Similarity 26.9%; Pred. No. 0.0022;
Matches   58; Conservative 27; Mismatches 92; Indels 39; Gaps 9;

Qy   33 VYWNSANKRFQAEGGYVILYPQIGDRLDLLCPRARPGPHSSPN--YEFYKLYLVGGAQGR   90
Db   28 VFWNSSNILLRKE-GVTLQVNVNDYLDIYCPHY------NSSQRGIAEQYVLMVSYRGYR   81

Qy   91 RCEAPPAPNLLLTCDR-----PDLDLRFTIKFQEYSPNLWGHEFRSHHDYYIIATSDGTREG 147
Db   82 TCD-PQLGFKRWECNRPHAPHAPIKFSEKFQRYSAPSLGYEFHVGGOEYYYISTP------ 134

Qy 148 LESLQGGVCLIRGMKVLLRVGQSPRGGAVPRKPVSEMPMERDRGAAHSLEPGKENLPGDP 207
Db 135 -THHGRSCLRLRVVYCCSTRASDSDDEPQPTEP----------DYTLRP---NIKIDD 178

Qy 208 TSNATSRGAEGPLPPPSMPAVAGAAGGLALLLLGVA 243
Db 179 LDD-----YDNPEVPKLEKSISGSSPSRDRLLLTVA 209

RESULT 44
ID   P91274   PRELIMINARY;   PRT;   305 AA.
AC   P91274;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Hypothetical 30.8 kDa protein.
GN   F26B1.4.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditina;
OC   Rhabditoidea; Pelodorinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   Latreille P., Wamsley P., Kramer J.;
RT   "The sequence of C. elegans cosmid F26B1.";
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RT   "Direct Submission.";
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U80444; AAB37788.1; -.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR002486; Col_cuticle_N.
DR   Pfam; PF01391; Collagen; 2.
DR   Pfam; PF01484; Col_cuticle_N; 1.
KW   Hypothetical protein.
SQ   SEQUENCE 305 AA; 30770 MW; 6BDC3ADD3E90A358 CRC64;

```
Qy  114 TIKFQEYSPNLWGHEFRSHHDYYIIATSDG------------------TREGLESLQGGVCLTRGMKVL 164
        ||||| ||||||||||||||||||||                        |||| | |||| ||||||
Db   47 TQHFQETSEKIW--------------DDLVFVATGTDRHKRHAYLTRAPVPVTGG-------------- 89

Qy  165 LRVGQSPRGGAVPRKPVSEMPMERDRGA------AHSL-----EPGKENLPG-------------- 205
        |||||||||||||||||||||||||||       |||       | ||||||
Db   90 FQTGRTANRGCNCGPGPGPGPGFNGEDGCNRFYCNLSHNLPQLPDDGBPEGVPGSSGGGDDIS 149

Qy  206 ------DP----TSNATSRGAEGPLPPPSMPAVAGAAGGIALLLLGVAGAGGAMCWRRRAK 257
              ||    ||  ||| | ||||||||||||||||||||||||||||
Db  150 VLRHDPGRCAQCPAGPMGPPGPPGPPGVYPGITGAPG-----LDGIPGRHGTPGY------ 198

Qy  258 PSESRHPGP-GSFGRGGSLGLGGGGGM-----------GPREAEPGELGIALRGGGAADPPF 307
                |  ||| || | |  |              | :|:  |  |:| :
Db  199 PGEPGLPGPGVGDPGPDGPDGKPGNPGNNGVRVRRIPGPGPGRGAE-GDSGV--QGG----PG 250

Qy  308 CPHYEKVSGDYGHPVYIVQDGPPQSPP 334
           ||  |  :  |||| :|| |||
Db  251 SPGLAGAPGDSGAP--------GPQGPP 270
```

RESULT 45
Q9XUE9

```
ID   Q9XUE9        PRELIMINARY;    PRT;    304 AA.
AC   Q9XUE9;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   F52B11.4 protein.
GN   F52B11.4.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditina;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RL   Matthews L.;
RL   Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99069613; PubMed=9851916;
RA   none;
RT   "Genome sequence of the nematode C.elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
DR   EMBL; Z82268; CAB55195.1; -.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR002486; Col_cuticle_N.
DR   Pfam; PF01391; Collagen; 2.
DR   Pfam; PF01484; Col_cuticle_N; 1.
SQ   SEQUENCE  304 AA;  29144 MW;  68C73E2251E72D8A CRC64;
```

```
Query Match     5.7%;   Score 140;  DB 5;  Length 304;
Best Local Similarity   26.8%;  Pred. No. 0.0035;
Matches  59;   Conservative   16;   Mismatches 87;  Indels  58;  Gaps  10;

Qy  142 DGTREGLESLQGGVCL--------------TRGMKVLLRVGQSPRGGAV--------------P 177
        || |||||||||||||              ||||||||||||||||||||            ||
Db   87 EGVDQGGSEGAQGGSCSGCCLPGAPAGTPG---KPGRPGRPGAGLPGNPGRPPAQP 141

Qy  178 RKPVSEMPM--------ERDRGAAHSLEP-GKENLPGDPTSNATSRGAEGPLPP-----PSMP 226
        :   :   :        | |:| ||| | :||| | |
Db  142 CEPITPPCKPCPQGPAGPAGPAGPQGPDAGAGPGAP-CQGSGAGAPGAPGAPGAPGNP 200

Qy  227 AVAGAAG--GIALLLLGVAGAGGAMCWRRRAKPSESRHP--------------GPGSPGRGGSLG 276
        |||||    ||||||||||||||||||||||||||                 |||| ||| ||
Db  201 GQAGAPGQPGADAQSESIPGAPG-------QAGPGQPGPAGSPGAPGSPGAPGGPGGPGAPGQKG 253

Qy  277 LGGGGGMGPREABPGELGIALRGGGAADPPPCPHYEKVSG 316
        |||||| :|: :|:  |  |:|  : : : : |  |  |
Db  254 PSGAPGQPGADGNPGAPGQPGQAGGAGEGKGICPKYCAIDG 293
```

RESULT 46
O97405

```
ID   O97405        PRELIMINARY;    PRT;    1378 AA.
AC   O97405;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Collagen pro alpha-chain precursor.
DE   HDCOL 1 ALPHA.
OS   Haliotis discus (Abalone).
OC   Eukaryota; Metazoa; Mollusca; Gastropoda; Archaeogastropoda;
OC   Haliotidae; Haliotis.
OX   NCBI_TaxID=36094;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=MUSCLE;
RX   MEDLINE=99234051; PubMed=10215888;
RA   Yoneda C., Hirayama Y., Nakaya M., Matsubara Y., Irie S., Hatae K.,
RA   Watabe S.;
RT   "The occurrence of two types of collagen proalpha-chain in the abalone
RT   Haliotis discus muscle.";
RL   Eur. J. Biochem. 261:714-721(1999).
DR   EMBL; AB017600; BAA75668.1; -.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001007; VWF_C.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 17.
DR   ProDom; PD000007; Collagen; 1.
DR   ProDom; PD000078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
KW   Signal.
FT   SIGNAL
SQ   SEQUENCE  1378 AA;  132583 MW;  6EEE34A65B52E634 CRC64;
```

```
Query Match     5.7%;   Score 140;  DB 5;  Length 1378;
Best Local Similarity   25.4%;  Pred. No. 0.02;
Matches  96;   Conservative  21;   Mismatches 125;  Indels 136;  Gaps  20;

Qy    2 GPPHSGPGGVR--VGALLLLGVLGVLVSGLSLEPVYWN--SANKRFQAEGGVVLYPQIGDR 57
        |||  ||| |    || :|||| |  |||  |||        | ::  |  || |||
Db  646 GPP--GPAGARGDAGAQGLIGMPG---ERGPIGRNGPQGNRGLTGERG----QDGE- 692

Qy   58 LDLLCPRARP-----------PGPHSSPNYEFVKLYLVGGAQGRRCEAPAPNLLLTCDR 106
         | |||| ||           || | ::|  | :| :|||  |  ||  || ::|
Db  693 ----PGRPGEAGARCGSPGQPGPSG----LVSAKGDRGEAGPA------------ 726

Qy  107 PDLDLRFTIKFQEYSPNLWGHEFRSHHDYYIIATSDGTR-EGLESLQGGVCLTRGMKVL 164
         |   :|  :|  | |||| | |:  |||| :|||| || ||||||||||||||||||
Db  727 --------GEPGP-------PGSSGQRGPAGAQGPQGPTGLSGPTGEM 759

Qy  165 LRVGQSPRGGAVPRKPVSEMPMERDRGAAHSLE------------PGKENLPGDP--T 208
        |||||||||||||||||||||||||||||||                 | ||||
Db  760 GQTGSDGKDGAKGDTGARGYPGEAGPIGAPGNEGREGRKGSRGSGGIPGNSGTGPDPGRA 819

Qy  209 SNATSRGAEGPLPPPSMPAVAGAAGGIALLLLGVAGAGGAMCWRRRAKPSESRHPG--- 265
        :||| | | | :|||||||||||||||||||||||||||||||||||||||||:|
Db  820 GPPGSPGAQGP------PGPSGAT------GLSGDGGE---RGETGPPGPRSGEEGAPG 862

Qy  266 ----PGSFGRGGSLGLGGGGG----MGPR----EAEPGELGIALRGGGAADPPPCPHYE 312
Db  863 MPGLDGAQGBERGSPGLINGPSGPPGPVGPQGBRGANGPPGPSGAGEA----GAAGPPGSAGEP 918

Qy  313 KVSGDYGHPVYIVQDGPP 330
Db  919 GLRGDNGNDGAPGQAGPP 936
```

```
RESULT 47
ID    Q9YIB4
AC    Q9YIB4;
DT    01-MAY-1999 (TrEMBLrel. 10, Created)
DT    01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT    01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE    Alpha 1 type I collagen.
OS    Cynops pyrrhogaster (Japanese common newt).
OC    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC    Amphibia; Batrachia; Caudata; Salamandroidea; Salamandridae; Cynops.
OX    NCBI_TaxID=8330;
RN    [1]
RP    SEQUENCE FROM N.A.
RC    TISSUE=REGENERATE FORELIMBS;
RX    MEDLINE=99407244; PubMed=10474166;
RA    Asahina K., Obara M., Yoshizato K.;
RT    "Expression of genes of type I and type II collagen in the formation
RT    and development of the blastema of regenerating newt limb.";
RL    Dev. Dyn. 216:59-71(1999).
DR    EMBL; AB015438; BAA36973.1; -.
DR    InterPro; IPR000087; Collagen.
DR    InterPro; IPR000885; Fib_collagen_C.
DR    InterPro; IPR001007; VWF_C.
DR    Pfam; PF01410; COLFI; 1.
DR    Pfam; PF01391; Collagen; 18.
DR    ProDom; PD000007; Collagen; 1.
DR    ProDom; PD002078; Fib_collagen_C; 1.
DR    SMART; SM00038; COLFI; 1.
DR    SMART; SM00214; VWC; 1.
DR    PROSITE; PS01208; VWFC; UNKNOWN_1.
KW    Collagen.
SQ    SEQUENCE  1450 AA;  137564 MW;  ABF8A74841B87B7C CRC64;

Query Match              5.7%;  Score 139.5; DB 13;  Length 1450;
Best Local Similarity   25.7%;  Pred. No. 0.023;
Matches  103;  Conservative  24;  Mismatches 137;  Indels 137;  Gaps  24;

Qy      2  GPPHSGPGPGGVRVGALLLGVLGL--VSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLD  59
              :|:            |    ||  :    ||   |   |   |       |   |||:|
Db    552  GPP--GPSGARGQS---GVWGPPGPKGPAAGEP-----------------GKSGER--  584

Qy     60  LLCPRARPPGPHSSPNYEFYKLYLVGGAQGRRCEAPPAPNLLTLTCDRPDLDLRFTIKFQE  119
              |||           | |    |  |    : |  : ||| || | : ||  |
Db    585  --GVAGPPGATGAPGKDG-----PPGPS-GPSGERGEQGPAGSPGPQG  630

Qy    120  Y--SPNIWGHEFRSHHDYYIIATSDGTREGLESLQG--GVCLITRGMKVLLRVGQSPRGGA  175
                    |   :     :  : :     |  |       |    |: |  :  |   |  |
Db    631  LPGSPGAGEAGKPGEQG---APGDAGPGPSGPRGERGFPGERG------GQGPAGAQ  680

Qy    176  VPR----KPVSE----------------MPMER---------DRGAAHSLEPG  199
                     |                          |             |
Db    681  GPRGSPGSPGQNDGAKGEAGANAGAPGGRGPPGLQGMPGERGSAGMPGAKGDRGDA-----G  735

Qy    200  KENLPGDPTSNATSRGAEGPLPPPSMPAVAG--AAGGLALLLGVAGAGGAMCWRRRRAK  257
                     |     | |      |         ||  |:   |     |
Db    736  TKGADGAPGKDG-ARGLTGPIGPPGPSGAPGDKGEGGPS-----GPAGPTGA-----  784

Qy    258  PSESRHPG------------PGSFGRGGSLG----LGGGGGMGPREAEPGELGTAL  297
              |  : |                 :   | |     |  : |    |   |
Db    785  PGERGEBPGAPGPAGICGPPGADGQPGAKGEBGDGAGPKGDAGAGPAGPTGA-PGPAGNVG  843

Qy    298  RGG-----GAADPPFCPHYEKVSGDYGHPVYIVQDGPPQSP  333
              |           |: :        |  |  :   || |
Db    844  APGFKGTRGAAGPPGATGPFGAAGRLGPPGPSGNAGPPGPP  884
```

```
RESULT 48
ID    Q9UMG6
AC    Q9UMG6;
DT    01-MAY-2000 (TrEMBLrel. 13, Created)
DT    01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT    01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE    Collagen type IV a6 chain.
DE    COL4A6.
GN    COL4A6.
OS    Homo sapiens (Human).
OC    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX    NCBI_TaxID=9606;
RN    [1]
RP    SEQUENCE OF 1-1644 FROM N.A.
RX    MEDLINE=96299642; PubMed=8661006;
RA    Zhang X., Zhou J., Reeders S.T., Tryggvason K.;
RT    "Structure of the human type IV collagen COL4A6 gene, which is mutated
RT    in Alport syndrome-associated leiomyomatosis.";
RL    Genomics 33:473-479(1996).
RN    [2]
RP    SEQUENCE FROM N.A.
RA    Zhang X.;
RL    Submitted (JAN-1996) to the EMBL/GenBank/DDBJ databases.
DR    EMBL; U47004; AAB19039.1; -.
DR    EMBL; U46960; AAB19039.1; JOINED.
DR    EMBL; U46961; AAB19039.1; JOINED.
DR    EMBL; U46962; AAB19039.1; JOINED.
DR    EMBL; U46963; AAB19039.1; JOINED.
DR    EMBL; U46964; AAB19039.1; JOINED.
DR    EMBL; U46965; AAB19039.1; JOINED.
DR    EMBL; U46966; AAB19039.1; JOINED.
DR    EMBL; U46967; AAB19039.1; JOINED.
DR    EMBL; U46968; AAB19039.1; JOINED.
DR    EMBL; U46969; AAB19039.1; JOINED.
DR    EMBL; U46970; AAB19039.1; JOINED.
DR    EMBL; U46971; AAB19039.1; JOINED.
DR    EMBL; U46973; AAB19039.1; JOINED.
DR    EMBL; U46974; AAB19039.1; JOINED.
DR    EMBL; U46975; AAB19039.1; JOINED.
DR    EMBL; U46976; AAB19039.1; JOINED.
DR    EMBL; U46977; AAB19039.1; JOINED.
DR    EMBL; U46979; AAB19039.1; JOINED.
DR    EMBL; U46980; AAB19039.1; JOINED.
DR    EMBL; U46981; AAB19039.1; JOINED.
DR    EMBL; U46982; AAB19039.1; JOINED.
DR    EMBL; U46983; AAB19039.1; JOINED.
DR    EMBL; U46984; AAB19039.1; JOINED.
DR    EMBL; U46985; AAB19039.1; JOINED.
DR    EMBL; U46986; AAB19039.1; JOINED.
DR    EMBL; U46987; AAB19039.1; JOINED.
DR    EMBL; U46988; AAB19039.1; JOINED.
DR    EMBL; U46989; AAB19039.1; JOINED.
DR    EMBL; U46990; AAB19039.1; JOINED.
DR    EMBL; U46991; AAB19039.1; JOINED.
DR    EMBL; U46992; AAB19039.1; JOINED.
DR    EMBL; U46993; AAB19039.1; JOINED.
DR    EMBL; U46994; AAB19039.1; JOINED.
DR    EMBL; U46995; AAB19039.1; JOINED.
DR    EMBL; U46996; AAB19039.1; JOINED.
DR    EMBL; U46997; AAB19039.1; JOINED.
DR    EMBL; U46998; AAB19039.1; JOINED.
DR    EMBL; U46999; AAB19039.1; JOINED.
DR    EMBL; U47000; AAB19039.1; JOINED.
DR    EMBL; U47001; AAB19039.1; JOINED.
DR    EMBL; U47002; AAB19039.1; JOINED.
DR    EMBL; U47003; AAB19039.1; JOINED.
DR    InterPro; IPR000087; Collagen.
DR    InterPro; IPR001442; ProcollagnC4.
DR    Pfam; PF01391; Collagen; 23.
DR    Pfam; PF01413; C4; 2.
DR    ProDom; PD000007; Collagen; 4.
DR    ProDom; PD003923; ProcollagnC4; 1.
DR    SMART; SM00111; C4; 2.
KW    Collagen.
SQ    SEQUENCE  1690 AA;  163696 MW;  4698AE2CC2D3E859 CRC64;
```

RESULT 49

```
Query Match    5.7%;  Score 139.5;  DB 4;  Length 1690;
Best Local Similarity  25.2%;  Pred. No. 0.027;
Matches  61;  Conservative  25;  Mismatches  83;  Indels  73;  Gaps  11;

Qy  147  GLESLQG--GVCLTRGMKVLL----RVGQSPRGGA----------------------VPRKPVSEM  184
                 :::|:|                                             |::|  :
Db  905  GFPGIPGLPGISGTRGLKGIPGSTGKMGPSGRAGTPGEKGDRGNPGPVGIPSPRRPMSNL  964

Qy  185  PMERDRGAAHSL------------------EPGKENLPGDPTSNATSRGAEG-PLPP-----  222
             |::|:|                  :|        |::                |:
Db  965  WLKGDKGSQGSAGSNGFPGPGRDKGEAGRPGPPGLPGAPGLPGIIKGVSGKPGPPGFMGI  1024

Qy  223  ------------------PSMPAVAGAAG----GL--ALLLLGVAGAGGAMCWRRRAKPSE  260
                           |:  ::| |       |:   :  |:  |:|
Db  1025 RGLPGLKGSSGITGFPGNPGESGSQGIRGSPGLPGASGLPGLKGDNGQTV--EISGSPGP  1082

Qy  261  SRHPGPGSF-GRGGSLGLGGGGMGPREAEPGELGTALRGGGAADPPFCPHYEKVSGDYG  319
             ::|  |  ::|:|| |  |  |:|                |:  |::|
Db  1083 KGQPGESGPFKGTKGRDGLIGNIGPPGPKKGEDGKVGVS----GDVGLPGAPGPPGVAGMRG  1138

Qy  320  HP  321
         :|
Db  1139 EP  1140
```

```
ID   Q9Y4L4;    PRELIMINARY;    PRT;   1691 AA.
AC   Q9Y4L4;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Collagen type IV a6 chain.
GN   COL4A6.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE OF 1-1645 FROM N.A.
RX   MEDLINE=96299642; PubMed=8661006;
RA   Zhang X., Zhou J., Reeders S.T., Tryggvason K.;
RT   "Structure of the human type IV collagen COL4A6 gene, which is mutated
RT   in Alport syndrome-associated leiomyomatosis.";
RL   Genomics 33:473-479(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Zhang X.;
RL   Submitted (JAN-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U47004; AAB19038.1; JOINED.
DR   EMBL; U46959; AAB19038.1; JOINED.
DR   EMBL; U46960; AAB19038.1; JOINED.
DR   EMBL; U46961; AAB19038.1; JOINED.
DR   EMBL; U46962; AAB19038.1; JOINED.
DR   EMBL; U46963; AAB19038.1; JOINED.
DR   EMBL; U46964; AAB19038.1; JOINED.
DR   EMBL; U46965; AAB19038.1; JOINED.
DR   EMBL; U46966; AAB19038.1; JOINED.
DR   EMBL; U46967; AAB19038.1; JOINED.
DR   EMBL; U46968; AAB19038.1; JOINED.
DR   EMBL; U46969; AAB19038.1; JOINED.
DR   EMBL; U46970; AAB19038.1; JOINED.
DR   EMBL; U46971; AAB19038.1; JOINED.
DR   EMBL; U46972; AAB19038.1; JOINED.
DR   EMBL; U46973; AAB19038.1; JOINED.
DR   EMBL; U46974; AAB19038.1; JOINED.
DR   EMBL; U46975; AAB19038.1; JOINED.
DR   EMBL; U46976; AAB19038.1; JOINED.
DR   EMBL; U46977; AAB19038.1; JOINED.
DR   EMBL; U46978; AAB19038.1; JOINED.
DR   EMBL; U46979; AAB19038.1; JOINED.
DR   EMBL; U46980; AAB19038.1; JOINED.
DR   EMBL; U46981; AAB19038.1; JOINED.
DR   EMBL; U46982; AAB19038.1; JOINED.
DR   EMBL; U46983; AAB19038.1; JOINED.
DR   EMBL; U46984; AAB19038.1; JOINED.
DR   EMBL; U46985; AAB19038.1; JOINED.
DR   EMBL; U46986; AAB19038.1; JOINED.
DR   EMBL; U46987; AAB19038.1; JOINED.
DR   EMBL; U46988; AAB19038.1; JOINED.
DR   EMBL; U46989; AAB19038.1; JOINED.
DR   EMBL; U46990; AAB19038.1; JOINED.
DR   EMBL; U46991; AAB19038.1; JOINED.
DR   EMBL; U46992; AAB19038.1; JOINED.
DR   EMBL; U46993; AAB19038.1; JOINED.
DR   EMBL; U46994; AAB19038.1; JOINED.
DR   EMBL; U46995; AAB19038.1; JOINED.
DR   EMBL; U46996; AAB19038.1; JOINED.
DR   EMBL; U46997; AAB19038.1; JOINED.
DR   EMBL; U46998; AAB19038.1; JOINED.
DR   EMBL; U46999; AAB19038.1; JOINED.
DR   EMBL; U47000; AAB19038.1; JOINED.
DR   EMBL; U47001; AAB19038.1; JOINED.
DR   EMBL; U47002; AAB19038.1; JOINED.
DR   EMBL; U47003; AAB19038.1; JOINED.
DR   Pfam; PF01391; Collagen; 23.
DR   Pfam; PF01413; C4; 2.
DR   ProDom; PD000007; Collagen; 4.
DR   ProDom; PD000000?; Collagen; 2.
DR   ProDom; PD003923; Procollagnc4; 1.
DR   SMART; SM00111; C4; 2.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR001442; ProcollagnC4.
KW   Collagen.
SQ   SEQUENCE  1691 AA;  163873 MW;  769A53DD1C7CA87 CRC64;
```

RESULT 50

```
ID   Q8TEJ5;    PRELIMINARY;    PRT;   705 AA.
AC   Q8TEJ5;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   FLJ00201 protein (Fragment).
GN   FLJ00201.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=SPLEEN;
```

```
Query Match    5.7%;  Score 139.5;  DB 4;  Length 1691;
Best Local Similarity  25.2%;  Pred. No. 0.027;
Matches  61;  Conservative  25;  Mismatches  83;  Indels  73;  Gaps  11;

Qy  147  GLESLQG--GVCLTRGMKVLL----RVGQSPRGGA----------------------VPRKPVSEM  184
                 :::|:|                                             |::|  :
Db  906  GFPGIPGLPGISGTRGLKGIPGSTGKMGPSGRAGTPGEKGDRGNPGPVGIPSPRRPMSNL  965

Qy  185  PMERDRGAAHSL------------------EPGKENLPGDPTSNATSRGAEG-PLPP-----  222
             |::|:|                  :|        |::                |:
Db  966  WLKGDKGSQGSAGSNGFPGPGRDKGEAGRPGPPGLPGAPGLPGIIKGVSGKPGPPGFMGI  1025

Qy  223  ------------------PSMPAVAGAAG----GL--ALLLLGVAGAGGAMCWRRRAKPSE  260
                           |:  ::| |       |:   :  |:  |:|
Db  1026 RGLPGLKGSSGITGFPGMPGESGSQGIRGSPGLPGASGLPGLKGDNGQTV--EISGSPGP  1083

Qy  261  SRHPGPGSF-GRGGSLGLGGGGMGPREAEPGELGTALRGGGAADPPFCPHYEKVSGDYG  319
             ::|  |  ::|:|| |  |  |:|                |:  |::|
Db  1084 KGQPGESGPFKGTKGRDGLIGNIGPPGPKKGEDGKVGVS----GDVGLPGAPGPPGVAGMRG  1139

Qy  320  HP  321
         :|
Db  1140 EP  1141
```

```
RA   Jikuya H., Takano J., Nomura N., Kikuno R., Nagase T., Ohara O.;
RT   "The nucleotide sequence of a long cDNA clone isolated from human
RT   spleen.";
RL   Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK074129; BAB84955.1; -.
NON_TER   1
FT   NON_TER   1
SQ   SEQUENCE   705 AA;   67430 MW;   07DB85A65A948BD3 CRC64;

Query Match             5.7%;   Score 139;   DB 4;   Length 705;
Best Local Similarity   25.8%;   Pred. No. 0.011;
Matches   73;   Conservative   18;   Mismatches   88;   Indels   104;   Gaps   14;

Qy  147  GLESLQG--------GVCLTRGMKVLLRVGQSPRGGAVP--------RKPVSEMPMERDRGAA  193
              ||  |                |   ||                    :|  |:
Db  289  GLPGPGGPSGAKGEPGTRGPPGLI----GPTGYGMPGLPGPKGDRGPAGVPGLLGDRG--  342

Qy  194  HSLEPGKENLPGD-----------PTSNAT--------SRGAEGPLPPSMPAVAGAAG  233
              ||||||  ||:                  ::        |  :|:  | |:
Db  343  ---EPGEDGEPGEQGPQGLGGPPGLPGSAGLPGRRGPPGGPKGEAGPGGGPPGVPGIRGDQG  399

Qy  234  --GLALLLGVAGAGGAMCWRRRARAKPSESRHPGP-----------------GSFGRG  272
            |  ||    ||     ||                                  |: |:
Db  400  PSGLA-----GKPGVPG-------ERGLPGAHGPPGPTGPKGEPGFTGRPGGPGVAGALGQK  449

Qy  273  GSLGL----------GGGGGMGPR-------EAEPGELGIALRG---------GGAAD  304
              |||           :||:                  |: |:          |||
Db  450  GDLGLPGPQGPGLRGPSGIPGLQGPAGPIGPQGLPGLKGEPGLPGPPGEGRAGEPGTAGPTG  509

Qy  305  PPFCPHYEKVSGDYGHPVYIVQDGPPGQSPPNIYYTSISVLEWP  347
              |:|:                    |:||:     :|:
Db  510  PPGVPGSPGITGPPGPP---GPPGPPGAPGAFDETGIAGLHLP  549
```

Search completed: February 11, 2003, 12:04:54
Job time : 49.7673 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2003, 11:47:24 ; Search time 19.4591 Seconds
           (without alignments)
           969.814 Million cell updates/sec

Title: US-10-021-121-2
Perfect score: 2450
Sequence: 1 MGPPHSGPGGVRVGALLLLG............TTLLRQRASVEABAGQHGPL 455

Scoring table: BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 50 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1841	75.1	340	1	EFB3_HUMAN	Q15768 homo sapien
2	1771	72.3	340	1	EFB3_MOUSE	O35393 mus musculu
3	628.5	25.7	336	1	EFB2_MOUSE	P52800 mus musculu
4	623	25.4	346	1	EFB1_HUMAN	P98172 homo sapien
5	620.5	25.3	333	1	EFB2_HUMAN	P52799 homo sapien
6	619	25.3	334	1	EFB1_CHICK	O73612 gallus gall
7	617.5	25.2	332	1	EFB2_BRARE	O73874 brachydanio
8	604.5	24.7	345	1	EFB1_MOUSE	P52795 mus musculu
9	599.5	24.5	345	1	EFB1_RAT	P52796 rattus norv
10	591	24.1	327	1	EFB1_XENLA	O13097 xenopus lae
11	185	7.6	345	1	EFA1_XENLA	P52794 xenopus lae
11	179	7.3	216	1	EFA1_BRARE	P79728 brachydanio
12	176	7.2	228	1	EFA2_BRARE	P52798 brachydanio
13	175.5	7.2	209	1	EFA4_HUMAN	O43921 homo sapien
14	172	7.0	213	1	EFA5_CHICK	P52804 gallus gall
15	170.5	7.0	228	1	EFA4_CHICK	P52802 gallus gall
16	170.5	7.0	201	1	EFA2_CHICK	P52801 gallus gall
17	169.5	6.9	228	1	EFA5_HUMAN	P52797 homo sapien
18	169.5	6.9	228	1	EFA5_BRARE	P79727 brachydanio
19	169.5	6.9	228	1	EFA5_HUMAN	P52797 homo sapien
20	167.5	6.8	228	1	EFA5_MOUSE	O08543 mus musculu
21	167.5	6.8	228	1	EFA1_HUMAN	P52803 homo sapien
22	167.5	6.8	228	1	EFA5_RAT	O08543 mus musculu
23	167.5	6.8	228	1	EFA2_HUMAN	P20827 homo sapien
24	166	6.8	205	1	EFA1_RAT	P97605 rattus norv
25	162.5	6.6	205	1	EFA1_RAT	P97553 rattus norv
26	161	6.6	206	1	EFA4_MOUSE	P97557 rattus nov
27	160.5	6.6	680	1	CA13_BOVIN	P04258 bos taurus
28	159	6.5	301	1	CA1A_MOUSE	Q05306 mus musculu
29	154.5	6.3	1049	1	CCO2_CAEEL	P17656 caenorhabdi
30	151.5	6.2	1670	1	CA34_HUMAN	Q01955 homo sapien
31	148.5	6.1	1527	1	CA3A_HUMAN	P39061 mus musculu
32	146	6.0	1027	1	CAFF_RIPPA	P30754 rifia pach
33	145	5.9	674	1	CAIA_BOVIN	P23206 bos taurus
34	144.5	5.9	1516	1	CA1H_HUMAN	P39060 homo sapien
35	144	5.9	675	1	CA39_CHICK	P32017 gallus gall
36	144	5.9	921	1	CA19_HUMAN	P20849 homo sapien
37	143	5.8	674	1	CA1A_CHICK	P08125 gallus gall
38	142.5	5.8	635	1	CA28_HUMAN	P25067 homo sapien
39	142.5	5.8	1496	1	CA25_HUMAN	P05997 homo sapien
40	142	5.8	744	1	CA18_HUMAN	P27658 homo sapien
41	142	5.8	1029	1	CA26_MOUSE	Q02788 mus musculu
42	142	5.8	1763	1	CA24_ASCSU	P27393 ascaris suu
43	141.5	5.8	1466	1	CA13_HUMAN	P02461 homo sapien
44	141	5.8	744	1	CA1B_RABIT	P14282 oryctolagus
45	140.5	5.7	743	1	CA19_MOUSE	Q00780 mus musculu
46	140	5.7	680	1	CA1A_HUMAN	Q03692 homo sapien
47	139.5	5.7	210	1	CA2_HAECO	P16252 haemonchus
48	139	5.7	1804	1	CA1A_MOUSE	Q61245 mus musculu
49	138.5	5.7	1459	1	CA2B_MOUSE	P28481 mus musculu
50	138.5	5.7	1736	1	CA2B_HUMAN	P13942 homo sapien

ALIGNMENTS

RESULT 1
EFB3_HUMAN

ID   EFB3_HUMAN   STANDARD;   PRT;   340 AA.
AC   Q15768; O00680; Q92875;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Ephrin-B3 precursor (EPH-related receptor tyrosine kinase ligand 8)
DE   (LERK-8) (EPH-related receptor transmembrane ligand ELK-L3).
GN   EFNB3 OR EPLG8 OR LERK8.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Cerretti D.P.;
RL   Submitted (JUL-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=97271551; PubMed=9126477;
RA   Tang X.X., Pleasure D.E., Ikegaki N.;
RT   "cDNA cloning, chromosomal localization, and expression pattern of
RT   Ephrin-B3, a new member of the EPLG gene family encoding ligands of EPH-
RT   related protein-tyrosine kinase receptors.";
RL   Genomics 41:17-24(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain cortex;
RX   MEDLINE=96404527; PubMed=8808709;
RA   Gale N.W., Flenniken A., Compton D.C., Jenkins N.A., Copeland N.G.,
RA   Gilbert D.J., Davis S., Wilkinson D.G., Yancopoulos G.D.;
RT   "Elk-L3, a novel transmembrane ligand for the Eph family of receptor
RT   tyrosine kinases, expressed in embryonic floor plate, roof plate and
RT   hindbrain segments.";
RL   Oncogene 13:1343-1352(1996).
CC   -!- FUNCTION: MAY PLAY A PIVOTAL ROLE IN FOREBRAIN FUNCTION. BINDS TO,
CC   AND INDUCE THE COLLAPSE OF, COMMISSURAL AXONS/GROWTH CONES IN
CC   VITRO. MAY PLAY A ROLE IN CONSTRAINING THE ORIENTATION OF
CC   LONGITUDINALLY PROJECTING AXONS (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN BRAIN; EXPRESSED IN
CC   EMBRYONIC FLOOR PLATE, ROOF PLATE AND HINDBRAIN SEGMENTS.
CC   -!- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.

CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; U57001; AAB05170.1; -.
DR   EMBL; U66406; AAC51203.1; -.
DR   EMBL; U62775; AAC50707.1; -.
DR   Genew; HGNC:3228; EFNB3.
DR   InterPro; IPR001799; Ephrin.
DR   MIM; 602297; -.
DR   Pfam; PF00812; Ephrin; 1.
DR   PRINTS; PR01347; EPHRIN.
DR   ProDom; PD002533; Ephrin; 1.
DR   PROSITE; PS01299; EPHRIN; 1.
KW   Developmental protein; Neurogenesis; Transmembrane; Glycoprotein;
KW   Signal; Polymorphism.
FT   SIGNAL   1   27
FT   CHAIN   28   340   EPHRIN-B3.
FT   DOMAIN   28   226   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   227   247   POTENTIAL.
FT   DOMAIN   248   340   CYTOPLASMIC (POTENTIAL).
FT   DOMAIN   338   340   PDZ RECOGNITION MOTIF (POTENTIAL).
FT   CARBOHYD   210   210   N-LINKED (GLCNAC...) (POTENTIAL).
FT   VARIANT   166   166   R -> Q.
FT   /FTId=VAR_002356.
SQ   SEQUENCE   340 AA; 35834 MW; EDF2A23C2FDE79F CRC64;

Query Match   75.1%; Score 1841; DB 1; Length 340;
Best Local Similarity   100.0%; Pred. No. 8.6e-115;
Matches 338; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 MGPPHSGPGGVRVGALILILGVLLGIVSGLSLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDL   60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MGPPHSGPGGVRVGALILILGVLLGIVSGLSLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDL   60

Qy   61 LCPRARPPGPHSSPNYEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDLDLRFTIKFQEY 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LCPRARPPGPHSSPNYEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDLDLRFTIKFQEY 120

Qy   121 SPNLWGHEFRSHHDYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKP 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 SPNLWGHEFRSHHDYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKP 180

Qy   181 VSEMPMERDRGAAHSLEPGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAGGLALLLL 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 VSEMPMERDRGAAHSLEPGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAGGLALLLL 240

Qy   241 GVAGAGGAMCWRRRRAKPSESSRHPGPGSFGRGGSLGLGGGGMGPREAEPGELGIALRGG 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GVAGAGGAMCWRRRRAKPSESSRHPGPGSFGRGGSLGLGGGGMGPREAEPGELGIALRGG 300

Qy   301 GAADPPFCHYEKVSGDYGHPVVIVQDGPPQSPPNIYY 338
           |||||||||||||||||||||||||||||||||||||
Db   301 GAADPPFCHYEKVSGDYGHPVVIVQDGPPQSPPNIYY 338

RESULT 2
EFB3_MOUSE   STANDARD;   PRT;   340 AA.
ID   EFB3_MOUSE
AC   O35593;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Ephrin-B3 precursor.
GN   EFNB3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=98143367; PubMed=9484836;
RA   Bergemann A.D., Zhang L., Chiang M.-K., Brambilla R., Klein R.,

---

RA   Flanagan J.G.;
RT   "Ephrin-B3, a ligand for the receptor EphB3, expressed at the midline
RT   of the developing neural tube.";
RL   Oncogene 16:471-480(1998).
RN   [2]
RP   FUNCTION.
RX   MEDLINE=20171264; PubMed=10704386;
RA   Imondi R., Wideman C., Kaprielian Z.;
RT   "Complementary expression of transmembrane ephrins and their receptors
RT   in the mouse spinal cord: a possible role in constraining the
RT   orientation of longitudinally projecting axons.";
RL   Development 127:1397-1410(2000).
CC   -!- FUNCTION: MAY PLAY A PIVOTAL ROLE IN FOREBRAIN FUNCTION. BINDS TO,
CC   AND INDUCE THE COLLAPSE OF, COMMISSURAL AXONS/GROWTH CONES IN
CC   VITRO. MAY PLAY A ROLE IN CONSTRAINING THE ORIENTATION OF
CC   LONGITUDINALLY PROJECTING AXONS.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: EXPRESSED ON LATERAL FLOOR PLATE CELLS,
CC   SPECIFICALLY ON COMMISSURAL AXON SEGMENTS THAT HAVE PASSED THROUGH
CC   THE FLOOR PLATE. EXPRESSED IN CELLS OF THE RETINAL GANGLION CELL
CC   LAYER DURING RETINAL AXON GUIDANCE TO THE OPTIC DISC.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED IN THE FLOOR PLATE THROUGHOUT THE
CC   PERIOD OF COMMISSURAL AXON PATHFINDING.
CC   -!- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; AF025288; AAC53537.1; -.
DR   MGD; MGI:109196; Efnb3.
DR   InterPro; IPR001799; Ephrin.
DR   Pfam; PF00812; Ephrin; 1.
DR   PRINTS; PR01347; EPHRIN.
DR   ProDom; PD002533; Ephrin; 1.
DR   PROSITE; PS01299; EPHRIN; 1.
KW   Developmental protein; Neurogenesis; Transmembrane; Glycoprotein;
KW   Signal.
FT   SIGNAL   1   27
FT   CHAIN   28   340   EPHRIN-B3.
FT   DOMAIN   28   227   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   228   248   POTENTIAL.
FT   DOMAIN   249   340   CYTOPLASMIC (POTENTIAL).
FT   DOMAIN   338   340   PDZ RECOGNITION MOTIF (POTENTIAL).
FT   CARBOHYD   210   210   N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   340 AA; 35884 MW; 52F3D58FD209A6B8 CRC64;

Query Match   72.3%; Score 1771; DB 1; Length 340;
Best Local Similarity   95.6%; Pred. No. 3.6e-110;
Matches 323; Conservative   7; Mismatches   8; Indels   0; Gaps   0;

Qy   1 MGPPHSGPGGVRVGALILILGVLLGIVSGLSLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDL   60
           ||||| ||||| ||||| ||| |||:|||||||||||||||||||||||||||||||||||||
Db   1 MGAPHFGPGGVQVGALLLLGFAGLVSGLSLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDL   60

Qy   61 LCPRARPPGPHSSPNYEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDLDLRFTIKFQEY 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LCPRARPPGPHSSPSYEFYKLYLVEGAQGRRCEAPPAPNLLLTCDRPDLDLRFTIKFQEY 120

Qy   121 SPNLWGHEFRSHHDYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKP 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 SPNLWGHEFRSHHDYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKP 180

Qy   181 VSEMPMERDRGAAHSLEPGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAGGLALLLL 240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||:||||||
Db   181 VSEMPMERDRGAAHSAEPGRDTIPGDPSSNATSRGAEGPLPPPSMPAVAGAAGGMALLLL 240

Qy   241 GVAGAGGAMCWRRRRAKPSESSRHPGPGSFGRGGSLGLGGGGMGPREAEPGELGIALRGG 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GVAGAGGAMCWRRRRAKPSESSRHPGPGSFGRGGSLGLGGGGMGPREAEPGELGIALRGG 300

Db   241 GVAGAGGAMCWRRRRAKPSESRHPGPGSFPGRGGSLGLGGGGMGPREAEPGELGIALRGG 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   301 GAADPPPCPHYEKVSGDYGHPVYIVQDGPPQSPPNIY 338

Db   301 GTADPPPCPHYEKVSGDYGHPVIIVQDGPPQSPPNIY 338

RESULT 3
ID   EFB2_MOUSE   STANDARD;   PRT;   336 AA.
AC   P52800;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Ephrin-B2 precursor (EPH-related receptor tyrosine kinase ligand 5)
DE   (LERK-5) (HTK ligand) (HTK-L) (ELF-2).
GN   EFNB2 OR EPLG5 OR HTKL OR ELF2 OR EPL5.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96145238; PubMed=8559144;
RA   Cerretti D.P., Vanden Bos T., Nelson N., Kozlosky C.J., Reddy P.,
RA   Maraskovsky E., Park L.S., Lyman S.D., Copeland N.G.,
RA   Gilbert D.J., Jenkins N.A., Fletcher R.A.;
RT   "Isolation of LERK-5: a ligand of the eph-related receptor tyrosine
RT   kinase.";
RL   Mol. Immunol. 32:1197-1205(1995).
SQ   SEQUENCE FROM N.A.

Qy   14 GALLLLGVLLGLVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDLLCPRARPPGPHSS 73
          |||||||||||||||||||||||||||||||||||||||||||||||||| | | ||
Db   17 GLLMVLCRTAISRSTVLEPIYWNSSNSKFLPGQGLVLYPQIGDKLDIICPKV---DSKTV 73

Qy   74 PNYEFYKLIVGGAQGRRCEAPPANPLLLTCDRPDLDLRFTIKFQEYSPNLWGHEFRSHH 133
          |||||| || | |:|:| ||||||||||||||||||||||||||||||||| || |:
Db   74 GQYEYYKVYMVDKDQADRCTIKKENTPLLNCARPDQDVKFTIKFQEFSPNIWGLEFQKNK 133

Qy   134 DYYIIATSDGTREGLESLQGGVCLITRGMKVLLRVGQ--SPRGGAVPRKEVSEMPMER-DR 190
          ||| || || || || | | :|| | | :| || ||||||||||||| : ||
Db   134 DYYIISTSNGSLEGLDNQEGGVCQTRAMKILMKVGQDASSAGSARNHGPTRRPELEAGTN 193

Qy   191 GAAHSLEPGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAGGLALLLLGVGAGGAMC 250
          ||||:|| | | | : |:|| || | | | | | ||| |||||||||||||||
Db   194 GRSSTTSPFVKPNPGSSTDGNSAGHSGNNLLGSEVALFAGIASGCIIFIVIILVLLL 253

Qy   251 WRRRRAKPSESRHPGPGSFPGRGGSLGLGGGGMGPREAEPGELGIALRGGGAADPPPFCPH 310
          | : : | | || |||||||| || ||||||||||||||||| || |||| |||
Db   254 KYRRRHRKHSPQHTTTLSLSTLAITPKRGGNN----NGSEPSDVIIPLR---TADSVFCPH 306

Qy   311 YEKVSGDYGHPVYIVQDGPPQSPPNIY 338
          |||||||||||| ||||:||||||||| |
Db   307 YEKVSGDYGHPVIIVQEMPPQSPANIY 334

RESULT 4
ID   EFB1_HUMAN   STANDARD;   PRT;   346 AA.
AC   P98172;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Ephrin-B1 precursor (EPH-related receptor tyrosine kinase ligand 2)
DE   (LERK-2) (ELK ligand) (ELK-L).
GN   EFNB1 OR EPLG2 OR LERK2 OR EFL-3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.

CC   ---
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---

DR   EMBL; U16819; AAA99708.1; -.
DR   EMBL; L38847; AAC42052.1; -.
DR   EMBL; U30244; AAA82934.1; -.
KW   Signal; Phosphorylation.
DE   Developmental protein; Neurogenesis; Transmembrane; Glycoprotein;
FT   SIGNAL   1   28
FT   CHAIN   29   336   EPHRIN-B2.
FT   DOMAIN   29   232   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM 233   253   POTENTIAL.
FT   DOMAIN   254   336   CYTOPLASMIC (POTENTIAL).
FT   DOMAIN   334   336   PDZ RECOGNITION MOTIF (POTENTIAL).
FT   CARBOHYD   39   39   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD 142   142   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CONFLICT   3   4   MISSING (IN REF. 3).
FT   CONFLICT 177   177   A -> T (IN REF. 1).
SQ   SEQUENCE   336 AA; 37202 MW; D088949986399554 CRC64;

Query Match   25.7%; Score 628.5; DB 1; Length 336;
Best Local Similarity 41.8%;
Matches 137; Conservative 49; Mismatches 129; Indels 13; Gaps 5;
Pred. No. 8.8e-35;

FUNCTION.
RL   [4]
Mol. Cell. Biol. 15:4921-4929(1995).
RP   SEQUENCE FROM N.A.
RC   STRAIN=ICR; TISSUE=Brain;
RX   MEDLINE=20171264; PubMed=10704386;
RA   Imondi R., Wideman C., Kaprielian Z.;
RT   "Complementary expression of transmembrane ephrins and their receptors
RT   in the mouse spinal cord: a possible role in constraining the
RT   orientation of longitudinally projecting axons.";
RL   Development 127:1397-1410(2000).
CC   -!- FUNCTION: MAY PLAY A ROLE IN CONSTRAINING THE ORIENTATION OF
CC   LONGITUDINALLY PROJECTING AXONS.
CC   -!- SUBUNIT: BINDS TO THE RECEPTOR TYROSINE KINASE EPHB4.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: EXPRESSED ON LATERAL FLOOR PLATE CELLS,
CC   SPECIFICALLY ON COMMISSURAL AXON SEGMENTS THAT HAVE PASSED THROUGH
CC   THE FLOOR PLATE. EXPRESSED IN CELLS OF THE RETINAL GANGLION CELL
CC   LAYER DURING RETINAL AXON GUIDANCE TO THE OPTIC DISC.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED IN THE FLOOR PLATE THROUGHOUT THE
CC   PERIOD OF COMMISSURAL AXON PATHFINDING.
CC   -!- PTM: INDUCIBLE PHOSPHORYLATION OF TYROSINE RESIDUES IN THE
CC   CYTOPLASMIC DOMAIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its

---

Qy   301 GAADPPPCPHYEKVSGDYGHPVYIVQDGPPQSPPNIY 338
Db   301 GTADPPPCPHYEKVSGDYGHPVIIVQDGPPQSPPNIY 338

RP   SEQUENCE FROM N.A.

RESULT 3
ID   EFB2_MOUSE   STANDARD;   PRT;   336 AA.
AC   P52800;
Bennett B.D., Zeigler F.C., Gu Q., Fendly B., Goddard A.D.,
Gillett N., Matthews W.;
"Molecular cloning of a ligand for the EPH-related receptor protein-
tyrosine kinase Htk.";
Proc. Natl. Acad. Sci. U.S.A. 92:1866-1870(1995).
[3]
SEQUENCE FROM N.A.
STRAIN=CB57BL/6J X SJL/J;
MEDLINE=95199254; PubMed=7534404;
Bergemann A.D., Cheng H.J., Brambilla R., Klein R., Flanagan J.G.;
"ELF-2, a new member of the Eph ligand family, is segmentally
expressed in mouse embryos in the region of the hindbrain and newly
forming somites.";
Mol. Cell. Biol. 15:4921-4929(1995).

```
RC   TISSUE=Placenta;
RX   MEDLINE=9439923; PubMed=8070404;
RA   Beckmann M.P., Cerretti D.P., Baum P., Vanderbos T., James L.,
RA   Farrah T., Kozlosky C., Hollingsworth T., Shilling H., Maraskovsky E.,
RA   Fletcher F.A., Lhotak V., Pawson T., Lyman S.D.;
RT   "Molecular characterization of a family of ligands for eph-related
RT   tyrosine kinase receptors.";
RL   EMBO J. 13:3757-3762(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Davis S., Gale N.W., Aldrich T.H., Maisonpierre P.C., Lhotak V.,
RA   Pawson T., Goldfarb M., Yancopoulos G.D.;
RL   Submitted (NOV-1994) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Fletcher F.A., Huebner K., Shaffer L.G., Monaco A., Mueller U.,
RA   Kozlosky C., Druck T., Simoneaux D.K., Fairweather N., Chelly J.,
RA   Cerretti D.P., Belmont J.W., Beckmann M.P., Lyman S.D.;
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Howden P.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: BINDS TO, AND INDUCE THE COLLAPSE OF, COMMISSURAL
CC       AXONS/GROWTH CONES IN VITRO. MAY PLAY A ROLE IN CONSTRAINING THE
CC       ORIENTATION OF LONGITUDINALLY PROJECTING AXONS (BY SIMILARITY).
CC   -!- SUBUNIT: BINDS TO THE RECEPTOR TYROSINE KINASES EPHB1 AND EPHA1.
CC       BINDS GRIP1 AND GRIP2.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: HEART, PLACENTA, LUNG, LIVER, SKELETAL MUSCLE,
CC       KIDNEY, PANCREAS.
CC   -!- INDUCTION: BY TNF-ALPHA.
CC   -!- PTM: INDUCIBLE PHOSPHORYLATION OF TYROSINE RESIDUES IN THE
CC       CYTOPLASMIC DOMAIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.
CC   -------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------------
DR   EMBL; U09304; AAA53093.1; -.
DR   EMBL; L37361; AAA52269.1; -.
DR   EMBL; U09303; AAB41127.1; -.
DR   EMBL; AL136092; CAB86409.1; -.
DR   Genew; HGNC:3226; EFNB1.
DR   MIM; 300035; -.
DR   InterPro; IPR001799; Ephrin.
DR   Pfam; PF00812; Ephrin; 1.
DR   PRINTS; PR01347; EPHRIN.
DR   ProDom; PD002533; Ephrin; 1.
DR   PROSITE; PS01299; EPHRIN; 1.
KW   Developmental protein; Neurogenesis; Transmembrane; Glycoprotein;
KW   Signal; Phosphorylation.
FT   SIGNAL        1     24
FT   CHAIN        25    346       EPHRIN-B1.
FT   DOMAIN       25    237       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    238    258       POTENTIAL.
FT   DOMAIN      259    346       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      344    346       PDZ RECOGNITION MOTIF (POTENTIAL).
FT   CARBOHYD    139    139       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   346 AA;  38006 MW;  473DD2F1A5BF89DE CRC64;

Query Match             25.4%;  Score 623;  DB 1;  Length 346;
Best Local Similarity   39.2%;  Pred. No. 2.1e-34;
Matches  143;  Conservative  48;  Mismatches 116;  Indels 58;  Gaps 9;

Qy      8 PGGVRVGAILLGVLLGVLSGL-----SLEPVVWNSANKRFQAEGGYVLYPQIGDRLDLL 61
            :|:|:   ||::||||::|::        ||: :  :|: :::    ::||:  ::||:
Db      4 PGQRWLGKWLVAMVVWALCRLATPLAKNLEPVSWSSLNPKFLSGKGLIVYPKIGDKLDII 63
```

```
Qy     62 CPRARPPGPHSSPNYEFYKLYLIVGGAQCGRRCEAPPAPNLLLTCDRPDLDLRFTIKFQEYS 121
          ||||| |        :|::|||:| :: :|        :| :    |:| |   ||||:||:
Db     64 CPRAEAGRP-----YEYYKLYLVRPEBQAAACSTVLDPNVLVTCNRPEQEIRFTIKFQEFS 118

Qy    122 PNLWGHEFRSHHDYIYIATSDGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKPV 181
          || :| ||:.HHDY:YI:TS:G: EGLE: :GGVC TR MK::::VGQ P ::: P: :
Db    119 PNYMGLEFKKHHDYYIYISTSNGSLEGLENREGGVCRTRTMKIIMKVGQDPNAVTPEQLTT 178

Qy    182 SEMPMERDRGAAHSLE-PGKENLPGDPTSNATSRGAEGPLPPSMPAVAGAAGGLA---- 236
          :   :: ::  ::: : .|. ..: ::::.  :: :: ::|          |:||
Db    179 SRPSKEADNTVKWATQAPGSRGSLGDSDGKHETVNQEEKSGP------GASGGSSGDPD 231

Qy    237 ------LILLGVAGAGA--------------MCWRRRRAKPSESRHPGPGSFGRGGSSLGL 277
              |::||||:| |              ||.:                    :|:
Db    232 GFFNSKVALFAAVGAGCVIFLLIIIFLTVLLLKRKHRKHTQQ---------RAAALSL 282

Qy    278 ----GGGGGMGPREAEPGELGIALRGGGAADPPFCPHYEKVSGDYGHPVYIVQDGPPQSP 333
              ||||| |:: |:|   |:: || ||||||||||||||||||||||||:: :|:|
Db    283 STLASPKGGSGTAGTEPSDIIIPLR---TTENNYCPHYEKVSGDYGHPVYIVQEMPGSP 339

Qy    334 PNIYY 338
          |||:|
Db    340 ANYY 344
```

RESULT 5
EFB2_HUMAN

```
ID   EFB2_HUMAN     STANDARD;      PRT;   333 AA.
AC   P52799;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Ephrin-B2 precursor (EPH-related receptor tyrosine kinase ligand 5)
DE   (LERK-5) (HTK ligand) (HTK-L).
GN   EFNB2 OR EPLG5 OR LERK5 OR HTKL.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96145238; PubMed=8559144;
RA   Cerretti D.P., Vanden Bos T., Nelson N., Kozlosky C.J., Reddy P.,
RA   Maraskovsky E., Park L.S., Lyman S.D., Copeland N.G., Gilbert D.J.,
RA   Jenkins N.A., Fletcher F.A.;
RT   "Isolation of LERK-5: a ligand of the eph-related receptor tyrosine
RT   kinases.";
RL   Mol. Immunol. 32:1197-1205(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=95199254; PubMed=7534404;
RA   Bennett B.D., Zeigler F.C., Gu Q., Fendly B., Goddard A.D.,
RA   Gillett N., Matthews W.;
RT   "Molecular cloning of a ligand for the EPH-related receptor protein-
RT   tyrosine kinase Htk.";
RL   Proc. Natl. Acad. Sci. U.S.A. 92:1866-1870(1995).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98192220; PubMed=9533549;
RA   Vogt T., Stolz W., Welsh J., Jung B., Kerbel R.S., Kobayashi H.,
RA   Landthaler M., McClelland M.;
RT   "Overexpression of Lerk-5/Eplg5 messenger RNA: a novel marker for
RT   increased tumorigenicity and metastatic potential in human malignant
RT   melanomas.";
RL   Clin. Cancer Res. 4:791-797(1998).
CC   -!- FUNCTION: MAY PLAY A ROLE IN CONSTRAINING THE ORIENTATION OF
CC       LONGITUDINALLY PROJECTING AXONS (BY SIMILARITY).
CC   -!- SUBUNIT: BINDS TO THE RECEPTOR TYROSINE KINASES EPHB4 AND EPHA3.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: LUNG AND KIDNEY.
```

Left column

```
CC  -!- PTM: INDUCIBLE PHOSPHORYLATION OF TYROSINE RESIDUES IN THE
CC      CYTOPLASMIC DOMAIN (BY SIMILARITY).
CC  -!- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.
CC  -----------------------------------------------------------------------
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----------------------------------------------------------------------
DR  EMBL; U16797; AAA99707.1; -.
DR  EMBL; L38734; AAC41752.1; -.
DR  EMBL; U81262; AAD03786.1; -.
DR  Genew; HGNC:3227; EFNB2.
DR  MIM; 600527; -.
DR  InterPro; IPR001799; Ephrin.
DR  Pfam; PF00812; Ephrin; 1.
DR  PRINTS; PR01347; Ephrin.
DR  ProDom; PD002533; Ephrin; 1.
DR  PROSITE; PS01299; EPHRIN; 1.
KW  Developmental protein; Neurogenesis; Transmembrane; Glycoprotein;
KW  Signal; Phosphorylation.
FT  SIGNAL       1     27
FT  CHAIN       28    333       POTENTIAL.
FT  DOMAIN      28    229       EPHRIN-B2.
FT  DOMAIN     230    250       EXTRACELLULAR (POTENTIAL).
FT  TRANSMEM   251    333       POTENTIAL.
FT  DOMAIN     251    333       CYTOPLASMIC (POTENTIAL).
FT  DOMAIN     331    333       PDZ RECOGNITION MOTIF (POTENTIAL).
FT  CARBOHYD    36     36       N-LINKED (GLCNAC...) (POTENTIAL).
FT  CARBOHYD   139    139       N-LINKED (GLCNAC...) (POTENTIAL).
SQ  SEQUENCE   333 AA;  36923 MW;  6D9932A63626AEA CRC64;
```

```
Query Match
Best Local Similarity   25.3%;     Pred. No. 2.9e-34;
Matches 133;  Conservative 52;  Mismatches 130;  Indels 13;  Gaps 5;
                                   Score 620.5;  DB 1;  Length 333;

Qy   14  GALILLGVLGLVSGLSLEPVYWNSANKRFQAEGGYYLYPQIGDRLDLLCPRARPCGPHSS   73
         | :| |||| :||||||||||||: ::|| :|||  ||||||||:| :|||| |
Db   14  GVLMVLCRTAISKSIVLEPIYWNSSNSKFLPGQGLVLYPQIGDKLDIICPKV---DSKTV   70

Qy   74  PNYEFYKLIVLVGGAQGRRCEAPPAPNILLLTCDRPDLDLRFTIKFQEYSPNILWGHEFRSHH  133
         | :|  | | : |  | |||  :: | | | | ||| | |||||||||||||:|:|||: :|
Db   71  GQYEYYVYVVMVDKDQADRCTIKKENTPLLNCAKPDQDIKFTIKFQEFSPNLWGLEFPQKNK   130

Qy  134  DYYILIATSDGTREGLSSLQGGVCLTRGMKVLLRVGQ--SPRGGAVPRKPVSEMPVER-DR   190
         |||| :|| |: ||||  | |||| | | ||::|:|||   | ||:| | |||| |: :|
Db  131  DYYIISTSNGSLBGLDNQEGGVCQTRAMKILMKVGQDASSAGSTRNKDPTRPELEAGTN    190

Qy  191  GAAHSLEPGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAGGLALILLGVAGAGGAMC   250
         | :|:| | :|||| | :|| : : |  | |    :| | || |||||||||||| |
Db  191  GRSSTTSPFVKPNPGSSTDGNSAGHSGNNILGSEVALFAGIASGCIIFIVIIILVVLL     250

Qy  251  WRRRRAKPSESRHPGPGSFGRGGSLGLGGGGGMGPREAEPGELGIALRGGGAADPPFCPH   310
         : ||| |||:|  || |:|  || |||  ||:|   |   |||||||| |||||||:|||
Db  251  KYRRRHRKHSPGHTTTLSLSTLATPKRSGNN----NGSEPSDIIIPLR--TADSVFCPH    303

Qy  311  YEKVSGDYGHPVYIVQDGPPQSPPNIYY   338
         |||||||||||||||||: ||||| |||
Db  304  YEKVSGDYGHPVYIVQEMPPQSPANIYY   331
```

```
RESULT 6
EFB1_CHICK     STANDARD;   PRT;   334 AA.
AC  O73612;
DT  15-JUL-1999 (Rel. 38, Created)
DT  15-JUL-1999 (Rel. 38, Last sequence update)
DT  16-OCT-2001 (Rel. 40, Last annotation update)
DE  Ephrin-B1 precursor (CEK5 ligand) (CEL5-L).
GN  EFNB1.
OS  Gallus gallus (Chicken).
```

Right column

```
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC  Gallus.
OX  NCBI_TaxID=9031;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=97223524; PubMed=9070326;
RA  Holash J.A., Soans C., Chong L.D., Shao H., Dixit V.M.,
RA  Pasquale E.B.;
RT  "Reciprocal expression of the Eph receptor Cek5 and its ligand(s) in
RT  the early retina.";
RL  Dev. Biol. 182:256-269(1997).
CC  -!- SUBUNIT: BINDS TO THE RECEPTOR TYROSINE KINASE EPHB2.
CC  -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC  -!- PTM: INDUCIBLE PHOSPHORYLATION OF TYROSINE RESIDUES IN THE
CC      CYTOPLASMIC DOMAIN (BY SIMILARITY).
CC  -!- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.
CC  -----------------------------------------------------------------------
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----------------------------------------------------------------------
DR  EMBL; U72394; AAC07986.1; -.
DR  InterPro; IPR001799; Ephrin.
DR  Pfam; PF00812; Ephrin; 1.
DR  PRINTS; PR01347; Ephrin.
DR  ProDom; PD002533; Ephrin; 1.
DR  PROSITE; PS01299; EPHRIN; 1.
KW  Developmental protein; Neurogenesis; Transmembrane; Glycoprotein;
KW  Signal; Phosphorylation.
FT  SIGNAL       1     25
FT  CHAIN       26    334       POTENTIAL.
FT  DOMAIN      26    231       EPHRIN-B1.
FT  DOMAIN     232    252       EXTRACELLULAR (POTENTIAL).
FT  TRANSMEM   232    252       POTENTIAL.
FT  DOMAIN     253    334       CYTOPLASMIC (POTENTIAL).
FT  DOMAIN     332    334       PDZ RECOGNITION MOTIF (POTENTIAL).
FT  CARBOHYD   135    135       N-LINKED (GLCNAC...) (POTENTIAL).
SQ  SEQUENCE   334 AA;  36858 MW;  48AF5568ED56CD5 CRC64;
```

```
Query Match
Best Local Similarity   39.3%;     Pred. No. 3.7e-34;
Matches 144;  Conservative 50;  Mismatches 100;  Indels 72;  Gaps 13;
                                   Score 619;  DB 1;  Length 334;

Qy    8  PGGVR---VGALILLGVLGLVSGLSLEPVYWNSANKRFQAEGGYYLYPQIGDRLDLLCPRA   65
         | |:|     | :||||| ||||| | ||||  ||  |:  |  |  |:|| | :|||| |
Db    4  PRGGRWLLGVLLALCRLAAPLAKSLEPVSWSAGNPKFMSGKGLVIYPEIGDKDIICPKA    63

Qy   66  RPPGPHSSPNYEFYKLIVLVGGAQGRRCEAPPAPNILLLTCDRPDLDLRFTIKFQEYSPNLW  125
         |     || : |  | | | |  | |||    | ||| | | || | |||||||||:|||
Db   64  BPSKP------YDYYKLYLVKKDQADACSTVMDPNVLVTCNRPEQEIRFTIKFQEFSPNYM   118

Qy  126  GHEFRSHHDYYILIATSDGTREGLESSLQGGVCLTRGMKVLLRVGQSPRGGAVPRKPVSEMP  185
         |:||: : |||| :|| |: ||||  |||||| | ||||::|:|||| | ||:| | |||:|
Db  119  GLEFKRQQDYFITSTSNGTLDGLENREGGVCQTRSMKIVMKVGQDP-NAVIPEQLTISRP   177

Qy  186  MER-------DRGAAHSL----EPGKENLPGDPTSNA--TSRGAEGPLPPPSMPAVAGA   231
         ::|          :: |:    ||||||| | :|||    : | |||| |  :| |||
Db  178  SKEADNTVKLVTQSPRHKVPTVEEPGK---PGSVNQNGQETQGPSDGFI--SSKVAVPAA   232

Qy  232  AGG-------LALILLGVAGAGGAMCWRRRRAKPSESRH-----   272
         :||        | :||||| |||| | : || ||:|
Db  233  IGAGCVIFILIIIFIVVLLIKI------RKHRHRKHTQQRAAALSLSTLLASPKCSGNA   283

Qy  273  GSLGLGGGGGMGPREAEPGELGIALRGGGAADPPFCPHYEKVSGDYGHPVYIVQDGPPQS   332
         || |  ||:|   |   |||||||| ||||||:|||||||||||||||||||||: ||:
Db  284  GS------EPSDIIIPLR---TTENNYCPHYEKVSGDYGHPVYIVQEMPPQS          326

Qy  333  PPNIYY   338
         |||||
Db            338
```

RESULT 7

| Db | 327 PANIYY 332 |

```
ID   EFB2_BRARE     STANDARD;      PRT;   332 AA.
AC   O73874;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Ephrin-B2 precursor.
GN   EFNB2 OR EFNB2A.
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98484455; PubMed=9765210;
RA   Durbin L., Brennan C., Shiomi K., Cooke J., Barrios A.,
RA   Shanmugalingam S., Guthrie B., Lindberg R., Holder N.;
RT   "Eph signaling is required for segmentation and differentiation of
RT   the somites.";
RL   Genes Dev. 12:3096-3109(1998).
CC   -!- SUBUNIT: BINDS TO THE RECEPTOR TYROSINE KINASE EPHB4.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- PTM: INDUCIBLE PHOSPHORYLATION OF TYROSINE RESIDUES IN THE
CC       CYTOPLASMIC DOMAIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AJ004863; CAA06168.1; -.
DR   ZFIN; ZDB-GENE-990415-67; efnb2a.
DR   InterPro; IPR001799; Ephrin.
DR   Pfam; PF00812; Ephrin; 1.
DR   PRINTS; PR01347; EPHRIN.
DR   ProDom; PD002533; Ephrin; 1.
DR   PROSITE; PS01299; EPHRIN; 1.
KW   Developmental protein; Neurogenesis; Transmembrane; Glycoprotein;
KW   Signal; Phosphorylation.
FT   SIGNAL        1     24
FT   CHAIN        25    332       EPHRIN-B2.
FT   DOMAIN       25    225       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    226    246       POTENTIAL.
FT   DOMAIN      247    332       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      330    332       PDZ RECOGNITION MOTIF (POTENTIAL).
SQ   SEQUENCE   332 AA;  36724 MW;  189ED82372C71C8B CRC64;
```

Query Match   25.2%; Score 617.5; DB 1; Length 332;
Best Local Similarity   41.8%; Pred. No. 4.6e-34;
Matches 143; Conservative 54; Mismatches 106; Indels 39; Gaps 12;

```
Qy     14 GALLLLGVLGLVSGLLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDLLCPRARPGPHSS 73
          ||:||  :||:||| || |:|| ||| || |:| ::|: : |:::   ::|| : | | |
Db     11 GVLVIACKVNLSRALLLDSIYWNTNTKIVPGQGLVLYPQIGDKMDIVCPRVE---GGSM 67

Qy     74 PNYEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPLDLRRTIKFQEYSPNLWGHEFRSHH 133
          |||||| :|:|| |: ||   || ||||| |||| |:|||:|||| |||||| ||||::
Db     68 EGVEYYKLYWVPLEQLKSCQVTKADTPLLNCVKPDDQDVKFRTLKFQBEFSPNLWGLEFRGK 127

Qy    134 DYYIIATSDGTREGLESLQGGVCLLTRGMKVLLRVGQSPRGGAVPRK-PVSEMPMERDRGA 192
          ||||::|:|| |:|| | :| ::|| :|:| ||::|::| | | |:: ||:| | ||||
Db    128 DYYISTSNGTMEGLDNQEGGVCKTKSMKIIMKVGQNPSDPISPKDYPTSYPPKHPDLGG 187

Qy    193 AHS----------LEP------GKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAGAGGLALLLLGV 242
                     :|||
Db        ...
```

RESULT 8

| Db | 188 KDSKSNEVLKPDASPHGEDK--GDGNKSSSVIGSEVAL----FACIASASVIIIIIML 241 |

```
Qy    243 AGAGGAMCWRRRAKPSRSRHPGGSFG------RGGSLGLGGGGGMGPREARPGELGTA 296
          ::||  |:|  ||: |::||  | |:      :|| |:|: || |  |: :: :|:|
Db    242 VFL--LLKYRRHRKGHS-PQHATLSLSTLATPKRGGS----GGNNNG---SEPSDIIIP 291

Qy    297 LRGGGAAADPPFCPHYEKYSGDYGHPVYIVQODGPPQSPNIYY 338
          |:||  ||  | ||||::||:|:|:||:|:|||  | ||:|||
Db    292 LR----TADSVPCPHYEKYSGDYGHPVIVQEMPQSPANIYY 330
```

```
ID   EFB1_MOUSE     STANDARD;      PRT;   345 AA.
AC   P52795;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Ephrin-B1 precursor (EPH-related receptor tyrosine kinase ligand 2)
DE   (LERK-2) (ELK ligand) (ELK-L) (STRA1 protein) (CEK5 receptor ligand)
DE   (CEK5-L).
GN   EFNB1 OR EPLG2 OR LERK2 OR STRA1 OR EPL2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/Sv;
RX   MEDLINE=95203867; PubMed=7896266;
RA   Fletcher F.A., Renshaw B., Hollingsworth T., Baum P., Lyman S.D.,
RA   Jenkins N.A., Gilbert D.J., Copeland N.G., Davison B.L.;
RT   "Genomic organization and chromosomal localization of mouse Eplg2,
RT   a gene encoding a binding protein for the receptor tyrosine kinase
RT   elk.";
RL   Genomics 24:127-132(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95377533; PubMed=7649373;
RA   Bouillet P., Oulad-Abdelghani M., Vicaire S., Garnier J.M.,
RA   Schuhbaur B., Dolle P., Chambon P.;
RT   "Efficient cloning of cDNAs of retinoic acid-responsive genes in P19
RT   embryonal carcinoma cells and characterization of a novel mouse gene,
RT   Stra1 (mouse LERK-2/Eplg2).";
RL   Dev. Biol. 170:420-433(1995).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=95014510; PubMed=7929389;
RA   Shao H., Lou L., Pandey A., Pasquale E.B., Dixit V.M.;
RT   "cDNA cloning and characterization of a ligand for the Cek5 receptor
RT   protein-tyrosine kinase.";
RL   J. Biol. Chem. 269:26606-26609(1994).
RN   [4]
RP   FUNCTION.
RX   MEDLINE=20171264; PubMed=10704386;
RA   Imondi R., Wideman C., Kaprielian Z.;
RT   "Complementary expression of transmembrane ephrins and their receptors
RT   in the mouse spinal cord: a possible role in constraining the
RT   orientation of longitudinally projecting axons.";
RL   Development 127:1397-1410(2000).
CC   -!- FUNCTION: BINDS TO, AND INDUCE THE COLLAPSE OF, COMMISSURAL
CC       AXONS/GROWTH CONES IN VITRO. MAY PLAY A ROLE IN CONSTRAINING THE
CC       ORIENTATION OF LONGITUDINALLY PROJECTING AXONS.
CC   -!- SUBUNIT: BINDS TO THE RECEPTOR TYROSINE KINASES EPHB1 AND EPHA1.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: EXPRESSED ON LATERAL FLOOR PLATE CELLS,
CC       SPECIFICALLY ON COMMISSURAL AXON SEGMENTS THAT HAVE PASSED THROUGH
CC       THE FLOOR PLATE. EXPRESSED IN CELLS OF THE RETINAL GANGLION CELL
CC       LAYER DURING RETINAL AXON GUIDANCE TO THE OPTIC DISC.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED IN THE FLOOR PLATE THROUGHOUT THE
CC       PERIOD OF COMMISSURAL AXON PATHFINDING.
```

CC -!- PTM: INDUCIBLE PHOSPHORYLATION OF TYROSINE RESIDUES IN THE
CC   CYTOPLASMIC DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.
CC ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch)
CC ---
DR EMBL; U07602; AAC53247.1; -.
DR EMBL; U07598; AAC53247.1; JOINED.
DR EMBL; U07599; AAC53247.1; JOINED.
DR EMBL; U07600; AAC53247.1; JOINED.
DR EMBL; Z48781; CAA88695.1; -.
DR EMBL; U12983; AAA53231.1; -.
DR MGD; MGI:102708; Efnb1.
DR InterPro; IPR001799; Ephrin.
DR Pfam; PF00812; Ephrin; 1.
DR PRINTS; PR01347; EPHRIN.
DR ProDom; PD002533; Ephrin; 1.
DR PROSITE; PS01299; EPHRIN; 1.
KW Developmental protein; Neurogenesis; Transmembrane; Glycoprotein;
KW Signal; Phosphorylation.
FT SIGNAL   1   24
FT CHAIN   25   345   EPHRIN-B1.
FT DOMAIN   25   236   EXTRACELLULAR (POTENTIAL).
FT TRANSMEM   237   257   POTENTIAL.
FT DOMAIN   258   345   CYTOPLASMIC (POTENTIAL).
FT DOMAIN   343   345   PDZ RECOGNITION MOTIF (POTENTIAL).
FT CARBOHYD   139   139   N-LINKED (GLCNAC...) (POTENTIAL).
FT CONFLICT   90   90   S -> T (IN REF. 2).
SQ SEQUENCE   345 AA; 37859 MW; 8C96FD3DC5CBC405 CRC64;

Query Match   24.7%; Score 604.5; DB 1; Length 345;
Best Local Similarity   37.9%;
Matches 136; Conservative 51; Mismatches 107; Indels 65; Gaps 10;

Qy   15 ALILLGVLGLIVSGL--SLEPVYWNSANKFQAEGGYVLYPQIGDRLDLLCPRARPPGPHS   72
          :::::|:|:||::||| :|:||:||||||:|:||:::|:||:|||:||||||:|:||||
Db   15 AMVVLTLCRLATPLAKNLEPVSWSSLNPKFLSGKGLVIYPKIGDKLDIICPRAEAGRP--   72

Qy   73 SPNYEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDLDLRFTIKFQEYSPNLWGHEFRSH 132
          ||||||::|:|||||||||||||:|::|||||||||||||||:||:|||||||||||||
Db   73 ---YEYKLYLVRPEQAAACSTVLDPNVLVTCNKPHQEIRFTIKFQEFSPNYMGLEFKKY 129

Qy   133 HDYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKPVSEMPMERDRGA 192
          ||:::|||:|:|||||||::||||||::::||:|:||:::::::::::::||:||||||
Db   130 HDYYITSTSNGSLEGLENREGGVCRTRTMKIVMKVGQDP-NAVTPEQLTTSRPSKESDNT 188

Qy   193 AHSLEPGKENLPGDPTSNATSRGAEGP-------LPPSMPAVAGAAGG------LA 236
          :::|||||:|:|::|:|::|||:|||| |||||||||||:|| ||
Db   189 VKT------------------ATQAPGRGSQGDSDGHETVNQEEKSGPGAGGSGDSDSFFNSK 236

Qy   237 LILLGVAGAGGA---------MCWRRRAKPSESRHPGPGSFGRGGSLGL----GG 279
          |||||:||:|| :|| :| :|:::|:|:|||||||||| ||
Db   237 VALFAAVGAGCVIFLILIIFLTVLLLKLRKHRKHTQQ-----------RAAALSLSTLIASP 287

Qy   280 GGGMGPREAEPGELGIALRGGGAAADPPFCHYEKVSGDYGHPVYIVQDGPPGSPNIYY 338
          ::::|:::||:::::||:::|::::::::||||:|||||||||:||:|:||:|||||:
Db   288 KGGSGTPAGTEPSDIIIPLR---TTENNYCPHYEKVSGDYGHPVYIVQEMPPQSPANIYY 343

RESULT 9
EFB1_RAT
ID EFB1_RAT   STANDARD;   PRT;   345 AA.
AC P52796;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Ephrin-B1 precursor (EPH-related receptor tyrosine kinase ligand 2)

```
Db  237  VALFAAVGAGCVIFLLIIIFLTVLLLKLRKRHRKHTQQ------------RAAALSLSTLASP  287
                                                              :::  ::
Qy  280  GGGMGPREAEPQELGIALRGGGAAADPPFCPHYEKVSGDYGHPVVIVQDGPPQSPPNIIY  338
                                       |||||||||||||||||||  ::::: ::
Db  288  KGDSGTAGTEFSDIIIPLR---TTENNYCPHYEKVSGDYGHPVVIVQEMPPQSPANIIY  343
```

RESULT 10

```
ID   EFB1_XENLA     STANDARD;      PRT;   327 AA.
AC   O13097;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Ephrin-B1 precursor (EPH-related receptor tyrosine kinase ligand 2)
DE   (LERK-2) (ELK ligand) (ELK-L) (XLERK).
GN   EFNB1 OR EPLG2 OR LERK2.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97316777; PubMed=9174051;
RA   Jones T.L., Karavanova I., Chong L., Zhou R.P., Daar I.O.;
RT   "Identification of XLerk, an Eph family ligand regulated during
RT   mesoderm induction and neurogenesis in Xenopus laevis.";
RL   Oncogene 14:2159-2166(1997).
CC   -!- FUNCTION: MAY HAVE A ROLE IN THE DEVELOPING MESENCHYMAL AND
CC       NERVOUS TISSUE.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: IN THE ADULT, EXPRESSED AT LOW LEVELS IN MOST
CC       ADULT TISSUES WITH INCREASED LEVELS OBSERVED IN THE KIDNEY,
CC       OOCYTES, OVARY AND TESTIS.
CC   -!- PTM: INDUCIBLE PHOSPHORYLATION OF TYROSINE RESIDUES IN THE
CC       CYTOPLASMIC DOMAIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.
CC   --------------------------------------------------------------------------
```
```
DR   EMBL; U31427; AAC35995.1; -.
DR   InterPro; IPR001799; Ephrin.
DR   Pfam; PF00812; Ephrin; 1.
DR   PRINTS; PR01347; Ephrin.
DR   ProDom; PD002533; Ephrin; 1.
DR   PROSITE; PS01299; EPHRIN; 1.
KW   Developmental protein; Neurogenesis; Transmembrane; Glycoprotein;
KW   Signal; Phosphorylation.
FT   SIGNAL        1     20
FT   CHAIN        21    327       EPHRIN-B1.
FT   DOMAIN       21    225       EPHRIN-B1.
FT   DOMAIN      226    246       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    226    246       POTENTIAL.
FT   DOMAIN      247    327       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      325    327       PDZ RECOGNITION MOTIF (POTENTIAL).
FT   CARBOHYD    131    131       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    202    202       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   327 AA;  36621 MW;  71230CE7F6BE5974 CRC64;
```

```
Query Match    24.1%;  Score 591;  DB 1;  Length 327;
Best Local Similarity   39.5%;  Pred. No. 2.6e-32;
Matches  144;  Conservative   43;  Mismatches 100;  Indels  78;  Gaps  12;

Qy   10  GVR--VGALLLGVLGLVSGSLEPYWNSANKRPQAEGGYVLYPQIGDRLDLLCPRA--  65
             :::|||:||:|| ::: :| |||::||| |:|  |::|||||:|||||:|||:  ::
Db    3  GLRRLLGLLLVLVLYRLCSALGKNLEPVTWNSQNPRISGKGLVLYPEIGDRLDIICPKGLF  62
```

```
Qy   66  RPPGPHSSSPNYEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDLDLRFTIKFQEYSPNIW  125
             :   ||| :   |:::||:|   :::|  |:|  :|:||| || |||:||||:::||:|
Db   63  QP------------YEYYKLIYMVRRDQLBACSTVIDPNVLVTCNPQKEYERTIKFQEFFSPNYM  114
Qy  126  GHEFRSHHDYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVRKPVSEMP  185
Db  170  ----YEYKLIYMVRRDQLBACSTVIDPNVLVTCNPQKEYERTIKFQEFFSPNYM------  ...
Qy  186  MERDRGAAHSLEPGKENLPGDPPTSNATSRGA-EGPLPPP------SMPAVAGAAGLA  236
Db       ----TTRPSKE--ADNTGKIATFGPWNGPVQNPGKSDTNLSDKPTGRWGVDGFF  217
Qy  237  LLLLGVAGAGGMC-----WRRRRAKPSE------SRHPGPGSFGRGG  273
Qy  274  SLGLGGGGGMGPREAEPQELGIALRGGGAAADPPFCPHYEKVSGDYGHPVVIVQDGPPQSP  333
Db  278  S------------EPSDIIIPLR---TTENNYCPHYEKVSGDYGHPVVIVQEMPPQSP  320
Qy  334  PNIIY  338
Db  321  ANIIY  325
```

RESULT 11

```
ID   EFA2_BRARE     STANDARD;      PRT;   195 AA.
AC   P79727;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Ephrin-A2 precursor (EPH-related receptor tyrosine kinase ligand 6)
DE   (LERK-6) (ELF-1) (ZFEPHL3).
GN   EFNA2 OR EPLG6 OR LERK6.
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RX   MEDLINE=97195707; PubMed=9043080;
RA   Brennan C., Monschau B., Lindberg R., Guthrie B., Drescher U.,
RA   Bonhoeffer F., Holder N.;
RT   "Two Eph receptor tyrosine kinase ligands control axon growth and may
RT   be involved in the creation of the retinotectal map in the
RT   zebrafish.";
RL   Development 124:655-664(1997).
CC   -!- FUNCTION: CONTROL AXON GROWTH AND MAY BE INVOLVED IN THE CREATION
CC       OF THE RETINO-TECTAL MAP.
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC       (potential).
CC   -!- TISSUE SPECIFICITY: WIDESPREAD EXPRESSION IN THE EMBRYO.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED IN THE PRESUMPTIVE MIDBRAIN OF
CC       DEVELOPING EMBRYOS FROM THE SIX-SOMITE STAGE. BY 24 HOURS IT IS
CC       EXPRESSED THROUGHOUT THE MIDBRAIN INCLUDING THE REGION OF THE
CC       PRESUMPTIVE TECTUM. AT LATER STAGES IT IS EXPRESSED IN A GRADED
CC       FASHION THROUGHOUT THE TECTUM.
CC   -!- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.
CC   --------------------------------------------------------------------------
```
```
DR   EMBL; Y09668; CAA70863.1; -.
DR   ZFIN; ZDB-GENE-990415-66; efna2.
```

Left column

```
DR   InterPro; IPR001799; Ephrin.
DR   Pfam; PF00812; Ephrin; 1.
DR   PRINTS; PR01347; EPHRIN; 1.
DR   ProDom; PD002533; Ephrin; 1.
DR   PROSITE; PS01299; EPHRIN; 1.
KW   Developmental protein; Neurogenesis; Glycoprotein; GPI-anchor; Signal.
FT   SIGNAL        1     16       POTENTIAL.
FT   CHAIN        17    195       EPHRIN-A2.
FT   CARBOHYD     32     32       N-LINKED (GlcNAc...) (POTENTIAL).
SQ   SEQUENCE   195 AA;  22688 MW;  9EE284FEB61D0C42 CRC64;
```

```
Query Match             7.6%;
Best Local Similarity   29.9%;  Pred. No. 9.3e-06;
Matches   63;  Conservative  21;  Mismatches  71;  Indels  56;  Gaps   7;

Qy    33  VYWNSANKRFQAEGGYVLYPQIGDRLDLLCPRARPPGPHSSPNYEFYKLYLVGGAQGRRC  92
             ||||:|||||                         ||:||:
Db    29  VYWNSSNSRFW-QGEYTVAVSINDYLDVVCPYYESPQPHS--RMERYILFMVNHDGYLTC  85

Qy    93  EAPPAPNLLLTCDR---PDLDLRFTIKFQEYSPNLWGHEHRSHHDYYIIAISDGIREGLE 149
             :|                |:||  |  ||::|||| |::|:
Db    86  EHRMRGFKRWECNRPQSPDGGPLRFSEKFQLFTPFSLGFEFRPGHEYYISSPHPNHAGKP 145

Qy   150  SLQGGVCLTRGMKVLLRVGQSPRGGAVPRKPVSEMPERDRGAAHSLEPGKRENLPGDPTS 209
             :||                                        GYESPEPFLTD 169
Db   146  CLK-------LKVYV-----------------------------------KPTSS---- 186

Qy   210  NATSRGAEGPLPPPSMPAVAGAGGLALLLLL 240
             :|:||                |  LAVLML 186
Db   170  QSQRGBADGPC-----------
```

```
RESULT 12
EFA3_HUMAN
ID   EFA3_HUMAN     STANDARD;      PRT;   238 AA.
AC   P52797;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Ephrin-A3 precursor (EPH-related receptor tyrosine kinase ligand 3)
DE   (LERK-3) (EHK1 ligand) (EHK1-L).
GN   EFNA3 OR EPLG3 OR LERK3 OR EFL-2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95140419; PubMed=7838529;
RA   Kozlosky C.J., Maraskovsky E., McGrew J.T., Vandenbos T.,
RA   Teepe M., Lyman S.D., Srinivasan S., Fletcher F.A., Gayle R.B. III,
RA   Cerretti D.P., Beckmann M.P.;
RT   "Ligands for the receptor tyrosine kinases hek and elk: isolation of
RT   cDNAs encoding a family of proteins.";
RL   Oncogene 10:299-306(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95063919; PubMed=7973638;
RA   Davis S., Gale N.W., Aldrich T.H., Maisonpierre P.C., Lhotak V.,
RA   Pawson T., Goldfarb M., Yancopoulos G.D.;
RT   "Ligands for EPH-related receptor tyrosine kinases that require
RT   membrane attachment or clustering for activity.";
RL   Science 266:816-819(1994).
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN BRAIN, SKELETAL MUSCLE, SPLEEN,
CC       THYMUS, PROSTATE, TESTIS, OVARY, SMALL INTESTINE, AND PERIPHERAL
CC       BLOOD LEUKOCYTES.
CC   -!- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.
```

Right column

```
DR   EMBL; U14187; AAC50078.1; -.
DR   EMBL; L37360; AAA52268.1; -.
DR   Genew; HGNC:3223; EFNA3.
DR   MIM; 601381; -.
DR   InterPro; IPR001799; Ephrin.
DR   Pfam; PF00812; Ephrin; 1.
DR   PRINTS; PR01347; EPHRIN; 1.
DR   ProDom; PD002533; Ephrin; 1.
DR   PROSITE; PS01299; EPHRIN; 1.
KW   Glycoprotein; GPI-anchor; Signal.
FT   SIGNAL        1     22       POTENTIAL.
FT   CHAIN        23    238       EPHRIN-A3.
FT   CARBOHYD     38     38       N-LINKED (GlcNAc...) (POTENTIAL).
FT   CARBOHYD     67     67       N-LINKED (GlcNAc...) (POTENTIAL).
FT   CARBOHYD    100    100       N-LINKED (GlcNAc...) (POTENTIAL).
FT   CONFLICT     71     74       MISSING (IN REF. 2).
SQ   SEQUENCE   238 AA;  26350 MW;  8EFD6AE8FE33FDDA CRC64;
```

```
Query Match             7.3%;
Best Local Similarity   28.4%;  Pred. No. 2.9e-05;
Matches   65;  Conservative  24;  Mismatches  80;  Indels  60;  Gaps  12;

Qy     7  GPGGVRVGALLLLGVLGLVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDLLCP--R  64
             |||                    ||||| |:|:     ||:||:|||  || |
Db    24  GPG-----------------NRHAVYWNSSNQHLRRE-GYTVQVNVNDYLDIYCPHYN  67

Qy    65  ARPPGPHSSP----NYEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDL---DLRFTIKF 117
             :|                ||:  ||  |  ||  :|:|||| || | |   |:|| |:|
Db    68  SSGVGPGAGPGPGGGAEQYVLYMVSRNGYRTCNASQGFK-RWECNRPHAPHSPIKFSEKF 126

Qy   118  QEYSPNLWGHEFRSHHDYYIIAISDGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVP 177
             | :   |  ||  |::||||   : : |  |      |  |:|| ||    |
Db   127  QRYSAPSLGYEFHAGHEYYYISTPTHNLH-------WKCLR--MKVFVCCASTSHSG--- 174

Qy   178  RKPVSEMP--------------------MERDRGAAHSLE-------PGKENLP 204
             :||                         MERDRGAAHSLE       PGK NLP
Db   175  EKPVPTLPQFTMGPNVKINVLEDFEGENPQVPKLEKSISGTSPKREHLP 223
```

```
RESULT 13
EFA2_MOUSE
ID   EFA2_MOUSE     STANDARD;      PRT;   209 AA.
AC   P52801;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Ephrin-A2 precursor (EPH-related receptor tyrosine kinase ligand 6)
DE   (LERK-6) (ELF-1) (CEK7-ligand) (CEK7-L).
GN   EFNA2 OR EPLG6 OR LERK6 OR ELF1 OR EPL6.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95181289; PubMed=7876076;
RA   Shao H., Lou L., Pandey A., Verderame M.F., Siever D.A., Dixit V.M.;
RT   "cDNA cloning and characterization of a Cek7 receptor
RT   protein-tyrosine kinase ligand that is identical to the ligand
RT   (ELF-1) for the Mek-4 and Sek receptor protein-tyrosine kinases.";
RL   Cell 79:157-168(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95007776; PubMed=7522971;
RA   Cheng H.J., Flanagan J.G.;
RC   STRAIN=Swiss Webster; TISSUE=Brain;
RT   "Identification and cloning of ELF-1, a developmentally expressed
RT   ligand for the Mek4 and Sek receptor tyrosine kinases.";
```

RL   J. Biol. Chem. 270:3467-3470(1995).
CC   -!- SUBUNIT: BINDS TO THE RECEPTOR TYROSINE KINASES EPHA3, EPHA4 AND
CC   EPHA5.
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC   (Potential).
CC   -!- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.
DR   EMBL; U14941; AAA53636.1; -.
DR   EMBL; U14752; AAA68520.1; -.
DR   MGD; MGI:102707; Efna2.
DR   InterPro; IPR001799; Ephrin.
DR   Pfam; PF00812; Ephrin; 1.
DR   PRINTS; PR01347; Ephrin.
DR   ProDom; PD02533; Ephrin; 1.
DR   PROSITE; PS01299; EPHRIN; 1.
KW   Glycoprotein; GPI-anchor; Signal.
FT   SIGNAL   1   20   POTENTIAL.
FT   CHAIN   21   209   EPHRIN-A2.
FT   CARBOHYD   38   38   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   170   170   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   184   184   N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   209 AA; 23586 MW; F19975445F25B9ABC CRC64;

Query Match   7.2%; Score 176; DB 1; Length 209;
Best Local Similarity   29.3%; Pred. No. 3.9e-05;
Matches   58; Conservative   19; Mismatches   69; Indels   52; Gaps   7;

Qy   33 VYWNSANKRFQAE------GGYVLYPQIGDRLDLLCPRARPPGPHSSPNYEFYKLYLVGGA 87
Db   35 VYWNSNPRFQVSAVGDGGGYTVEVSINDYLDIYCPHYGAPLP-PAERMERYILYMVNGE 93

Qy   88 QGRRCEAPPAPNLLLTCDRPDL---DLRFTIKFQEYSPNLWGHEFRSHHDYYIIATSDGT 144
Db   94 GHASCDHRQGFKRWECNRPAAPGGPLKFSEKFQLFTPFSLGFEFRPGHEYYISATP-- 151

Qy   145 REGLESLQGGVCLTRGMKVLLRVGQSPRGAVPRKPVSEMPMERDRGAAHSLEPGKENLP 204
Db   152 ----PNLVDRPCLR--LKYVVR-------------------------PTNETLY 174

Qy   205 GDP----TSNATSRGAEG 218
Db   175 EAPEPIFTSNSSCSLGG 192

RESULT 14
ID   EPA2_HUMAN   STANDARD;   PRT;   213 AA.
AC   O43921; O76020;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Ephrin-A2 precursor (EPH-related receptor tyrosine kinase ligand 6)
DE   (LERK-6) (HEK7-ligand) (HEK7-L).
GN   EFNA2 OR EPLG6 OR LERK6.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98126446; PubMed=9465306;
RA   Ceretti D.P., Nelson N.;
RT   "Characterization of the genes for mouse LERK-3/Ephrin-A3 (Epl3),
RT   mouse LERK-4/Ephrin-A4 (Epl4), and human LERK-6/Ephrin-A2 (EPLG6);
RT   conservation of intron/exon structure.";

RL   Genomics 47:131-135(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Lamerdin J.E., McCready P.M., Skowronski E., Adamson A.W.,
RA   Burkhart-Schultz K., Gordon L., Kyle A., Ramirez M., Stilwagen S.,
RA   Phan H., Velasco N., Garnes J., Danganan L., Poundstone P.,
RA   Christensen M., Georgescu A., Avila J., Liu S., Attix C., Andreise T.,
RA   Trakhheim M., Amico-Keller G., Coefield J., Duarte S., Lucas S.,
RA   Bruce R., Thomas P., Quan G., Kronmiller B., Arellano A.,
RA   Montgomery M., Ow D., Nolan M., Trong S., Kobayashi A., Olsen A.O.,
RA   Carrano A.V.;
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RT   TISSUE=Brain;
RX   MEDLINE=99045414; PubMed=9826538;
RA   Aasheim H.C., Pedeutour F., Grosgeorge J., Logtenberg T.;
RT   "Cloning, chromosomal mapping, and tissue expression of the gene
RT   encoding the human Eph-family receptor tyrosine kinase ligand ephrin-A2.";
RL   Biochem. Biophys. Res. Commun. 252:378-382(1998).
CC   -!- SUBUNIT: BINDS TO THE RECEPTOR TYROSINE KINASES EPHA3, EPHA4 AND
CC   EPHA5.
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC   (Potential).
CC   -!- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.
DR   EMBL; U92896; AAC39577.1; -.
DR   EMBL; U92893; AAC39577.1; JOINED.
DR   EMBL; U92894; AAC39577.1; JOINED.
DR   EMBL; AC004258; AAC04996.1; -.
DR   EMBL; AJ007292; CAA07435.1; -.
DR   Genew; HGNC:3222; EFNA2.
DR   MIM; 602756; -.
DR   InterPro; IPR001799; Ephrin.
DR   Pfam; PF00812; Ephrin; 1.
DR   PRINTS; PR01347; Ephrin.
DR   ProDom; PD002533; Ephrin; 1.
DR   PROSITE; PS01299; EPHRIN; 1.
KW   Glycoprotein; GPI-anchor; Signal.
FT   SIGNAL   1   24   POTENTIAL.
FT   CHAIN   25   213   EPHRIN-A2.
FT   CARBOHYD   42   42   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   174   174   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   188   188   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CONFLICT   6   6   R -> A (IN REF. 3).
FT   CONFLICT   6   6   R -> A (IN REF. 3).
FT   CONFLICT   25   26   RA -> PP (IN REF. 3).
FT   CONFLICT   29   30   AA -> RR (IN REF. 3).
SQ   SEQUENCE   213 AA; 23878 MW; 33C9F81A8168B2D0 CRC64;

Query Match   7.2%; Score 175.5; DB 1; Length 213;
Best Local Similarity   36.8%; Pred. No. 4.3e-05;
Matches   43; Conservative   14; Mismatches   51; Indels   9; Gaps   3;

Qy   33 VYWNSANKRFQA------EGGYVLYPQIGDRLDLLCPRARPPGPHSSPNYEFYKLYLVGGA 87
Db   39 VYWNRSNPRFHAGAGDDGGGYTVEVSINDYLDIYCPHYGAPLP-PAERMEHVLVMVNGE 97

Qy   88 QGRRCEAPPAPNLLLTCDRPDL---DLRFTIKFQEYSPNLWGHEFRSHHDYYIIATS 141
Db   98 GHASCDHRQGFKRWECNRPAAPGGPLKFSEKFQLFTPFSLGFEFRPGHEYYISAT 154

RESULT 15
ID   EPA2_CHICK   STANDARD;   PRT;   200 AA.

AC   P52802;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Ephrin-A2 precursor (EPH-related receptor tyrosine kinase ligand 6)
DE   (LERK-6) (ELF-1).
GN   EFNA2 OR EPLG6 OR LERK6 OR ELF1.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95360981; PubMed=7634327;
RA   Cheng H.J., Nakamoto M., Bergemann A.D., Flanagan J.G.;
RT   "Complementary gradients in expression and binding of ELF-1 and Mek4
RT   in development of the topographic retinotectal projection map.";
RL   Cell 82:371-381(1995).
CC   -!- SUBUNIT: BINDS TO THE RECEPTOR TYROSINE KINASES EPHA3, EPHA4 AND
CC   EPHA5 (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC   (Potential).
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN A GRADIENT ACROSS THE TECTUM
CC   BEING MORE STRONGLY EXPRESSED AT THE POSTERIOR POLE.
CC   -!- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.

DR   EMBL; L40932; AAC42229.1; -.
DR   InterPro; IPR001799; Ephrin.
DR   Pfam; PF00812; Ephrin; 1.
DR   PRINTS; PR01347; EPHRIN.
DR   ProDom; PD002533; Ephrin; 1.
DR   PROSITE; PS01299; EPHRIN; 1.
KW   Glycoprotein; GPI-anchor; Signal.
FT   SIGNAL   1   22
FT   CHAIN   23   200   EPHRIN-A2.
FT   CARBOHYD   36   36   N-LINKED (GLCNAC...). (POTENTIAL).
FT   CARBOHYD   161   161   N-LINKED (GLCNAC...). (POTENTIAL).
FT   CARBOHYD   175   175   N-LINKED (GLCNAC...). (POTENTIAL).
SQ   SEQUENCE   200 AA; 23049 MW; 8FAB1AE5E45EED96 CRC64;

Query Match   35.3%;
Best Local Similarity   49%;
Matches 49; Conservative 16; Mismatches 58; Indels 16; Gaps 5;

Qy   15 ALLLLGVTLGLVSGLSLEP------------VYWNSANKRFQAEGGYVLYPQIGDRLDLLCPRA 65
            |||:|||:||||:|||| | ||||||||||||||||||||||||||||:|
Db   7 AALLAAIVG-VCVWSDDPGKVISDRYAVYWNRSNPRFH-RGDYTVEVSINDYLDIYCPHY 64

Qy   66 RPPGPHSSPNYEFYKLYLVGGAQGRCEAPAPNLLLTCDRPDL----DLRFTIKFQEYSP 122
            :||:|||||||||||||||||||||||||||||||||| ||:||:|||| |
Db   65 EEPLP--AERMERYVLIMVNYEGHASCDHRQKGFKRWECNRPDSPSGPLKFSEKFQLFTP 122

Qy 123 NLWGHEFRSHDYYIITATS 141
            ||:|||||:||||||:|:
Db 123 FSLGFEFRPGHEYYISAS 141

RESULT 16
EFA5_CHICK
AC   P52804;
ID   EFA5_CHICK   STANDARD;   PRT;   228 AA.
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)

DE   Ephrin-A5 precursor (EPH-related receptor tyrosine kinase ligand 7)
DE   (LERK-7) (RAGS protein).
GN   EFNA5 OR RAGS.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96360980; PubMed=7634326;
RA   Drescher U., Kremoser C., Handwerker C., Loschinger J., Noda M.,
RA   Bonhoeffer F.;
RT   "In vitro guidance of retinal ganglion cell axons by RAGS, a 25 kDa
RT   tectal protein related to ligands for Eph receptor tyrosine
RT   kinases.";
RL   Cell 82:359-370(1995).
CC   -!- FUNCTION: INDUCES GROWTH CONE COLLAPSE AND REPULSION OF RETINAL
CC   GANGLION CELL AXONS.
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC   (Potential).
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN A GRADED FASHION ACROSS THE
CC   TECTUM BEING MORE STRONGLY EXPRESSED TOWARDS THE POSTERIOR POLE.
CC   -!- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.

DR   EMBL; X90377; CAA62027.1; -.
DR   InterPro; IPR001799; Ephrin.
DR   Pfam; PF00812; Ephrin; 1.
DR   PRINTS; PR01347; EPHRIN.
DR   ProDom; PD002533; Ephrin; 1.
DR   PROSITE; PS01299; EPHRIN; 1.
KW   Developmental protein; Neurogenesis; Glycoprotein; GPI-anchor; Signal.
FT   SIGNAL   1   20
FT   CHAIN   21   228   EPHRIN-A5.
FT   CARBOHYD   37   37   N-LINKED (GLCNAC...). (POTENTIAL).
SQ   SEQUENCE   228 AA; 26206 MW; 56D8E4FBD8CF18AD CRC64;

Query Match   28.9%;
Best Local Similarity   73%;
Matches 73; Conservative 30; Mismatches 91; Indels 59; Gaps 13;

Qy   16 LLLLLGVLGL-VSGLSLEP------------VYWNSANKRFQAEGGYVLYPQIGDRLDLLCPR 64
            ||||:|:|| |||||| | ||||||||||||||||||||||||||||:|
Db   6 MLLLAVAALWVCVRGQEPGRKAVADRYAVYWNSTNPRFQ-QGDYHIDVCINDYLDVFCPH 64

Qy   65 ARPPGPHSSPNYEFYKLYLVG------------GAQGRCEAPAPNLLLTCDRPDLDLR 112
            :||:|||||||||||||| ||:||||||||||||||||||:|:|
Db   65 YEDSVPEDKT--ERVLYMVNFDGYSSCDHISKGFKRWECNRPHSPN------------GPLK 113

Qy 113 FTIKFQEYSPNLWGHEFRSHDYYIIATS----DGTREGLESLQGGVCLTRGMKVLLRVGQ 169
            |||:||::||:||:|||:|:||||:| ||:|:|||:|:||||||:::|::|:|
Db 114 FSEKFQLFTPFSLGFEFRPGREYFYISSAIPDNGRRS------------CLK--LKYFVR--- 159

Qy 170 SPRGGAVPRKPVSEMPERDRGAAHSLEPGXENLPGDPTSNATSRGAEGPLPPPSMP--A 227
            :|:::: :: :::|:||:: :|||:||||: |:|:| ::|:||:|:| |||:|
Db 160 -PANSCMKTIGVHDRVFDVNDKVENSLEPADDTV---RESAEPSRG-ENAAQTPRIPRL 214

Qy 228 VAGAAGGIALLLL 240
            :|:|:||:|||||
Db 215 LATLLFLLAMLLI 227

RESULT 17
EFA4_HUMAN

```
ID   EFA4_HUMAN     STANDARD;      PRT;    201 AA.
AC   P52798; O95457;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Ephrin-A4 precursor (EPH-related receptor tyrosine kinase ligand 4)
DE   (LERK-4).
GN   EFNA4 OR EPLG4 OR LERK4.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=B-cell;
RX   MEDLINE=95140419; PubMed=7838529;
RA   Kozlosky C.J., Maraskovsky E., McGrew J.T., Vandenbos T.,
RA   Teepe M., Lyman S.D., Srinivasan S., Fletcher F.A., Gayle R.B. III,
RA   Cerretti D.P., Beckmann M.P.;
RT   "Ligands for the receptor tyrosine kinases hek and elk: isolation of
RT   cDNAs encoding a family of proteins.";
RL   Oncogene 10:299-306(1995).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RX   MEDLINE=20076261; PubMed=10607706;
RA   Aasheim H.C., Munthe E., Funderud S., Smeland E.B., Beiske K.,
RA   Logtenberg T.;
RT   "A splice variant of human ephrin-A4 encodes a soluble molecule that
RT   is secreted by activated human B lymphocytes.";
RL   Blood 95:221-230(2000).
CC   -!- FUNCTION: MAY PLAY A ROLE IN THE INTERACTION BETWEEN ACTIVATED B
CC       LYMPHOCYTES AND DENDRITIC CELLS IN TONSILS.
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC       OR SECRETED (DEPENDING ON THE ISOFORM).
CC   -!- ALTERNATIVE PRODUCTS: 2 isoforms; 1/GPI-anchored (shown here) and
CC       2/secreted; are produced by alternative splicing.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN THE ADULT SPLEEN, LYMPH NODE,
CC       PROSTATE, OVARY, SMALL INTESTINE AND COLON, AND IN FETAL HEART,
CC       LUNG, LIVER AND KIDNEY. ALSO DETECTED IN HEMATOPOIETIC CELL LINES.
CC   -!- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.
```

```
DR   EMBL; U14188; AAC50079.1; -.
DR   EMBL; AJ006352; CAA06992.1; -.
DR   EMBL; AJ006353; CAA06993.1; -.
DR   Genew; HGNC:3224; EFNA4.
DR   MIM; 601380; -.
DR   InterPro; IPR001799; Ephrin.
DR   Pfam; PF00812; Ephrin; 1.
DR   PRINTS; PR01347; Ephrin.
DR   ProDom; PD002533; Ephrin; 1.
DR   PROSITE; PS01299; Ephrin; 1.
KW   Glycoprotein; GPI-anchor; Signal; Alternative splicing.
FT   SIGNAL        1     22
FT   CHAIN        23    170       EPHRIN-A4.
FT   PROPEP      171    201       REMOVED IN MATURE FORM (POTENTIAL).
FT   CARBOHYD     33     33       N-LINKED (GLCNAC...) (POTENTIAL).
FT   LIPID       170    170       GPI-ANCHOR (POTENTIAL).
FT   VARSPLIC    157    201       KSESAHPVGSPGESGTSGWRGGDTPSPLCLILLLLILILRL
FT                                LRIL -> NLPSHPKEPPSSQDPLEEBSSLLPALGVPIQTD
FT                                KME (IN ISOFORM 2).
SQ   SEQUENCE   201 AA;  22386 MW;  AB88D5443A9AF28D CRC64;
```

```
Query Match             6.9%;
Best Local Similarity   29.9%;
Matches  66;  Conservative  18;  Mismatches  82;  Indels  55;  Gaps  10;

Score 169.5;  DB 1;  Length 201;
Pred. No. 0.0001;

Qy    24  LVSGLSL-EPVYWNSANKRPQOAEGGYVLYPQIGDRLLDLLCPRARPPGPHSSPNYEFYKLY   82
              |||   ||||:|||:||   |:|||::|    |||:|:|::||  ||  |:||
Db    20  LRGGSSLRHVVVWNSSNPRL-LRGDAVVELGLNDYLDIVCPHYEGPGPGPEP--ETFALY   76

Qy    83  LVGGAQGRRCEAP-PAPNLLLTCDRPDLDLRFILKFQEYSSPNLWGHEFRSHHDYYIIATS  141
          |||| |||  :||  :  |:|||| |  | ||||||||||:|||:|| |::|| |:|||
Db    77  MVDWPGYESCQASGPRAYXRWVCSLPFGHVQFSEKIQRPTPFSLGFFLPGFTYYYSVP    136

Qy   142  DGTREGLESLQGGVCLTRGMKVLLRVGGQSPRGGAVPRKPVSEMPERDRGAAHSLEPGKE   201
           ::||::|  ::||::||:|    ||:  |   |:   |::| :  |   | :| ||
Db   137  --TPE-----SSGQCL------------------------RLQVSVCCKERKSESAHPV---  164

Qy   202  NLPGDPTSNATS--RGAEGPLPPPSMPAVAGAAGGLALLLL  240
          |||||||| :|:   ||   | |||:  |||  |||:|||
Db   165  ---GSPGESGTSGWRGGDTPSP------------LCLLLL  189
```

```
RESULT 18
ID   EFA5_BRARE     STANDARD;      PRT;    228 AA.
AC   P79728;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Ephrin-A5 precursor (EPH-related receptor tyrosine kinase ligand 7)
DE   (LERK-7) (AL-1) (ZFEPHL4).
GN   EFNA5 OR EFNA5B OR AL1 OR ELG7 OR LERK7.
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RX   MEDLINE=97195707; PubMed=9043080;
RA   Brennan C., Monschau B., Lindberg R., Guthrie B., Drescher U.,
RA   Bonhoeffer F., Holder N.;
RT   "Two Eph receptor tyrosine kinase ligands control axon growth and may
RT   be involved in the creation of the retinotectal map in the
RT   zebrafish.";
RL   Development 124:655-664(1997).
CC   -!- FUNCTION: CONTROL AXON GROWTH AND MAY BE INVOLVED IN THE CREATION
CC       OF THE RETINO-TECTAL MAP.
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC       (potential).
CC   -!- TISSUE SPECIFICITY: WIDESPREAD EXPRESSION IN THE EMBRYO.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED IN THE PRESUMPTIVE MIDBRAIN OF
CC       DEVELOPING EMBRYOS FROM THE SIX-SOMITE STAGE. BY 24 HOURS IT IS
CC       STRONGLY EXPRESSED IN THE MIDBRAIN CAUDAL TO THE PRESUMPTIVE
CC       TECTUM. AT LATER STAGES IT IS MAINTAINED AT THE POSTERIOR MARGIN
CC       OF THE TECTUM.
CC   -!- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.
```

```
DR   EMBL; Y09669; CAA70864.1; -.
DR   ZFIN; ZDB-GENE-980526-186; efna5b.
DR   InterPro; IPR001799; Ephrin.
DR   Pfam; PF00812; Ephrin; 1.
DR   PRINTS; PR01347; Ephrin.
DR   ProDom; PD002533; Ephrin; 1.
DR   PROSITE; PS01299; Ephrin; 1.
KW   Developmental protein; Neurogenesis; Glycoprotein; GPI-anchor; Signal.
FT   SIGNAL        1     20       POTENTIAL.
FT   CHAIN        21    228       EPHRIN-A5.
```

FT   CARBOHYD   37   37   N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   228 AA; 26995 MW; 74B34060C50541BB6E CRC64;
     Query Match
     Best Local Similarity   27.6%; Pred. No. 0.00012;
     Matches   72; Conservative   28; Mismatches   84; Indels   77; Gaps   13;

Qy   16 LLILGVLGLVSGLSLEP-----------------VYWNSANKRFQAEGGYVLYPQIGDRLDLLCPRA 65
          :||||:||:|||:|||:: |:|| ||| || || || | || || ||| ||
Db   7 IVFVGVILMWCVFSQEPSSKVMADRYAVFWNRTNPRFQ-RGDYHIDVCINDYLDIVCPHY 65

Qy   66 RPPGPHSSPNYEFYKLYLVG-----------------GAQGRRCEAPPAPNLLLTCDRPDLDLRF 113
          ||| :||| |||| ||||| | || | || | ||||| ||| ||||
Db   66 EDSVPEB--RTERVLIMYNYDGYSTCDHTAKGFXRWECNRPHSPN-----------GPLKF 114

Qy   114 TIKFQEYSPNLWGHEFRSHHDYYIIA---TSDGTREGLESLQGGVCLTRGMKVLLRVGQS 170
          ::|||||| ||||||||||||||| |||||||||||||||||| |||||| |||||
Db   115 SEKFQLFTPPSLGFEFPRGREYYYISSMTETGRRS-------CLK--LKVFVR---P 160

Qy   171 PRG-----GAVPRKPVSEM---PMERDRGAAHSLEPGKENLPGDPTSNATSRGAEGPLPP 222
          | | || || | | | || ||| | | | | | | | || || || |
Db   161 PNGCEKTIGVHDRVFVDDKVDNALEPRDTSHEAEPSRSDVSTSGLRHQTSR-----PL-- 214

Qy   223 PSMPAVAGAAGGLALLLLGVA 243
          ||||| | |||||| ||
Db   215 -------------LALLLCIS 223

RESULT 19

ID   EFA1_XENLA   STANDARD;   PRT;   216 AA.
AC   P52794;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Ephrin-A1 precursor (EPH-related receptor tyrosine kinase ligand 1)
DE   (LERK-1) (XELF).
GN   EFNA1 OR EPLG1 OR ELF.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97003306; PubMed=8843391;
RA   Weinstein D.C., Rahman S.M., Ruiz J.C., Hemmati-Brivanlou A.;
RT   "Embryonic expression of eph signalling factors in Xenopus.";
RL   Mech. Dev. 57:133-144(1996).
CC   -!- SUBUNIT: BINDS TO THE RECEPTOR TYROSINE KINASES EPHA2, EPHA4,
CC   EPHA5, EPHA6 AND EPHA7. ALSO BINDS WITH LOW AFFINITY TO EPHA1 (BY
CC   SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (By
CC   similarity).
CC   -!- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; U31204; AAA74485.1; -.
DR   EMBL; U31205; AAA74486.1; -.
DR   InterPro; IPR001799; Ephrin.
DR   Pfam; PF00812; Ephrin; 1.
DR   PRINTS; PR01347; EPHRIN.
DR   ProDom; PD005533; EPHRIN.
DR   PRODOM; PD005231; EPHRIN; 1.
DR   PROSITE; PS01299; EPHRIN; 1.
KW   Glycoprotein; GPI-anchor; Alternative splicing; Signal.
FT   SIGNAL   1   28

FT   CHAIN   29   216   EPHRIN-A1.
FT   CARBOHYD   36   36   N-LINKED (GLCNAC...) (POTENTIAL).
FT   VARSPLIC   162   216   TPPPVNVHTPRSHIQSDEPEVPLPGVMKSVAGNSAAPGTPC
FT   TLYGLLLAALLLRL -> SE (IN ISOFORM A').
SQ   SEQUENCE   216 AA; 24755 MW; 1B3A5O8E0A7B872E CRC64;
     Query Match
     Best Local Similarity   29.2%; Pred. No. 0.00015;
     Matches   71; Conservative   26; Mismatches   93; Indels   53; Gaps   12;

Qy   12 RVGALILLGVLGLVSGLSLEP---------VYWNSANKRFQAEGGYVLYPQIGDRLDLLCPR 64
          | || |||||:||:|||::|| |:|| ||| || || || | || || |||
Db   6 RAAVQLIVGV-GLGVGLWLREAQGERHIVFWNNSSNYRFMQE-DYTVQVLNDYLDIVCPY 63

Qy   65 ARPPGPHSSPNYEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDL---DLRFTIKFQEYS 121
          | ||| :||| |||| |||| || || | ||||| ||| || || ||||||
Db   64 YB--EGSVAGHTVERYLFLVDYEEYETCKPKSKDQVRWECNKPFAPHGPEKFCEKFQKFT 122

Qy   122 PNLWGHEFRSHHDYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKPV 181
          |||||||||||||||:| ||||||||||||||||| ||||||||||||| ||||||||
Db   123 PFTLGTEFREGRTYYIISKP-----IHYHGETCMR------LRVHVS---GRITPPPV 166

Qy   182 SEMPMERDRGAAHSLEPGKENLPGDPTSNATSRGAEGEGPLPPPSMPAVAG--AAGGLALL 239
           :| | | : :| | | : | | | || || || ||||||| ||||
Db   167 N----------VHT-------PRSHIQSDEPEVPL-PGVMKSVAGNSAAPGTPCTL 204

Qy   240 LGV 242
          ||
Db   205 YGL 207

RESULT 20

ID   EFA5_HUMAN   STANDARD;   PRT;   228 AA.
AC   P52803;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Ephrin-A5 precursor (EPH-related receptor tyrosine kinase ligand 7)
DE   (LERK-7) (AL-1).
GN   EFNA5 OR EPLG7 OR LERK7.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=97392664; PubMed=7748564;
RA   Winslow J.W., Moran P., Valverde J., Shih A., Yuan J.Q., Wong S.C.,
RA   Tsai S.P., Goddard A., Henzel W.J., Hefti F.;
RT   "Cloning of AL-1, a ligand for an Eph-related tyrosine kinase
RT   receptor involved in axon bundle formation.";
RL   Neuron 14:973-981(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95267434; PubMed=9245480;
RA   Kozlosky C.J., Vandenbos T., Park L.S., Cerretti D.P., Carpenter M.K.;
RT   "LERK-7: a ligand of the Eph-related kinases is developmentally
RT   regulated in the brain.";
RL   Cytokine 9:540-549(1997).
RN   [3]
RP   FUNCTION.
RX   MEDLINE=20069483; PubMed=10601038;
RA   Davy A., Gale N.W., Murray E.W., Klinghoffer R.A., Soriano P.,
RA   Feuerstein C., Robbins S.M.;
RT   "Compartmentalized signaling by GPI-anchored ephrin-A5 requires the
RT   Fyn tyrosine kinase to regulate cellular adhesion.";
RL   Genes Dev. 13:3125-3135(1999).
CC   -!- FUNCTION: MAY FUNCTION ACTIVELY TO STIMULATE AXON FASCICULATION.
CC   INDUCES COMPARTMENTALIZED SIGNALING WITHIN A CAVEOLAE-LIKE
CC   MEMBRANE MICRODOMAIN WHEN BOUND TO THE EXTRACELLULAR DOMAIN OF ITS
CC   COGNATE RECEPTOR. THIS SIGNALING EVENT REQUIRES THE ACTIVITY OF

Left column

```
CC       THE FYN TYROSINE KINASE.
RL   -!- SUBUNIT: BINDS TO THE RECEPTOR TYROSINE KINASES EPHA2, EPHA3 AND
CC       EPHB1.
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR IT
CC       IS COMPARTMENTALIZED IN DISCRETE CAVEOLAE-LIKE MEMBRANE
CC       MICRODOMAINS.
CC   -!- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.
CC   --------------------------------------------------------------------------
```

```
DR   EMBL; U26403; AAB60377.1; -.
DR   Genew; HGNC:3225; EFNA5.
DR   MIM; 601535; -.
DR   InterPro; IPR001799; Ephrin.
DR   Pfam; PF00812; Ephrin; 1.
DR   PRINTS; PR01347; EPHRIN.
DR   ProDom; PD002533; Ephrin; 1.
DR   PROSITE; PS01299; EPHRIN; 1.
KW   Developmental protein; Neurogenesis; Glycoprotein; GPI-anchor; Signal;
KW   Polymorphism.
FT   SIGNAL       1     20       POTENTIAL.
FT   CHAIN       21    228       EPHRIN-A5.
FT   CARBOHYD    37     37       N-LINKED (GLCNAC...) (POTENTIAL).
FT   VARIANT     55     55       N -> K (IN DBSNP:469062).
FT                               /FTId=VAR_012035.
SQ   SEQUENCE   228 AA;  26297 MW;  6893B1CCACFF3F57 CRC64;
```

```
Query Match      6.8%;  Score 167.5;  DB 1;  Length 228;
Best Local Similarity   28.8%;  Pred. No. 0.00016;
Matches   65;  Conservative  29;  Mismatches  81;  Indels  51;  Gaps  11;

Qy   33  VYWNSANKRFQAEGGYVLYPQIGDRLDLLCPRARPPGPHSSPNYEFYKLYLVG------   85
             ||||| ||||  ||| |||| | | |||||||| ||| |     | || ||
Db   34  VYWNSSNPRFQ-RGDYHIDVCINDYLDVFCPHYEDSVPEDKT--ERYVLYMVNFDGYSAC  90

Qy   86  -----GAQGRRCEAPPAPNLLLTCDRPDLDLRPTIKFQEYSPNLWGHEFRSHHDYYIIAT  140
               |  :  ||| :|| ||| ||| | | |||| ||| ||  :| :||| | ||:
Db   91  DHTSKGFRWECNRPHSPN--------GPLKFSEKFQLFTPFSLGFEFRPGREYFYISS   141

Qy  141  S---DGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKPVSEMPMERDRGAAHSLE  197
         :   ||| :  |   || ||| : |:|| |       || ||          | || |||
Db  142  AIPDNGRRS-------------CLK--LKVFVR------PTNSCMKIIGVHDRVFDVNDKVENSLE  186

Qy  198  PGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAGGLALILLGVA   243
         |.|  || | :|||:|  |: ||  :          |:|:|||| |
Db  187  PADDTV---HESAEPSRG-ENAAQTPIPSRL-----LAILLFLLA   223
```

```
RESULT 21
ID   EFA5_MOUSE    STANDARD;      PRT;    228 AA.
AC   O08543; O08544;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Ephrin-A5 precursor (EPH-related receptor tyrosine kinase ligand 7)
DE   (LERK-7) (AL-1).
GN   EFNA5 OR EPLG7 OR EPL7.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97060019; PubMed=8903354;
RA   Flenniken A.M., Gale N.W., Yancopoulos G.D., Wilkinson D.G.;
RT   "Distinct and overlapping expression patterns of ligands for
```

Right column

```
RT   Eph-related receptor tyrosine kinases during mouse embryogenesis.";
RL   Dev. Biol. 179:382-401(1996).
CC   -!- FUNCTION: MAY FUNCTION ACTIVELY TO STIMULATE AXON FASCICULATION.
CC       INDUCES COMPARTMENTALIZED SIGNALING WITHIN A CAVEOLAE-LIKE
CC       MEMBRANE MICRODOMAIN WHEN BOUND TO THE EXTRACELLULAR DOMAIN OF ITS
CC       COGNATE RECEPTOR. THIS SIGNALING EVENT REQUIRES THE ACTIVITY OF
CC       THE FYN TYROSINE KINASE (BY SIMILARITY).
CC   -!- SUBUNIT: BINDS TO THE RECEPTOR TYROSINE KINASES EPHA2, EPHA3 AND
CC       EPHB1 (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR IT
CC       IS COMPARTMENTALIZED IN DISCRETE CAVEOLAE-LIKE MEMBRANE
CC       MICRODOMAINS (BY SIMILARITY).
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC       SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.
CC   --------------------------------------------------------------------------
```

```
DR   EMBL; U90664; AAB50239.1; -.
DR   EMBL; U90665; AAB50240.1; -.
DR   MGD; MGI:1074444; Efna5.
DR   InterPro; IPR001799; Ephrin.
DR   Pfam; PF00812; Ephrin; 1.
DR   PRINTS; PR01347; EPHRIN.
DR   ProDom; PD002533; Ephrin; 1.
DR   PROSITE; PS01299; EPHRIN; 1.
KW   Developmental protein; Neurogenesis; Glycoprotein; GPI-anchor; Signal;
KW   Alternative splicing.
FT   SIGNAL       1     20       POTENTIAL.
FT   CHAIN       21    228       EPHRIN-A5.
FT   CARBOHYD    37     37       N-LINKED (GLCNAC...) (POTENTIAL).
FT   VARSPLIC   163    189       MISSING (IN SHORT ISOFORM).
SQ   SEQUENCE   228 AA;  26339 MW;  85439F5337420022 CRC64;
```

```
Query Match      6.8%;  Score 167.5;  DB 1;  Length 228;
Best Local Similarity   28.8%;  Pred. No. 0.00016;
Matches   65;  Conservative  29;  Mismatches  81;  Indels  51;  Gaps  11;

Qy   33  VYWNSANKRFQAEGGYVLYPQIGDRLDLLCPRARPPGPHSSPNYEFYKLYLVG------   85
             ||||| ||||  ||| |||| | | |||||||| ||| |     | || ||
Db   34  VYWNSSNPRFQ-RGDYHIDVCINDYLDVFCPHYEDSVPEDKT--ERYVLYMVNFDGYSAC  90

Qy   86  -----GAQGRRCEAPPAPNLLLTCDRPDLDLRPTIKFQEYSPNLWGHEFRSHHDYYIIAT  140
               |  :  ||| :|| ||| ||| | | |||| ||| ||  :| :||| | ||:
Db   91  DHTSKGFRWECNRPHSPN--------GPLKFSEKFQLFTPFSLGFEFRPGREYFYISS   141

Qy  141  S---DGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKPVSEMPMERDRGAAHSLE  197
         :   ||| :  |   || ||| : |:|| |       || ||          | || |||
Db  142  AIPDNGRRS-------------CLK--LKVFVR------PTNSCMKIIGVHDRVFDVNDKVENSLE  186

Qy  198  PGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAGGLALILLGVA   243
         |.|  || | :|||:|  |: ||  :          |:|:|||| |
Db  187  PADDTV---HESAEPSRG-ENAAQTPIPSRL-----LAILLFLLA   223
```

```
RESULT 22
ID   EFA5_RAT      STANDARD;      PRT;    228 AA.
AC   P97605;
DT   01-NOV-1997 (Rel. 35, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Ephrin-A5 precursor (EPH-related receptor tyrosine kinase ligand 7)
DE   (LERK-7) (AL-1).
GN   EFNA5 OR EPLG7 OR LERK7.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=Sprague-Dawley;
RX   MEDLINE=95267434; PubMed=7748564;
RA   Winslow J.W., Moran P., Valverde J., Shih A., Yuan J.Q., Wong S.C.,
RA   Tsai S.P., Goddard A., Henzel W.J., Hefti F., Beck K.D., Caras I.W.;
RT   "Cloning of AL-1, a ligand for an Eph-related tyrosine kinase
RT   receptor involved in axon bundle formation.";
RL   Neuron 14:973-981(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley;
RA   Li Y.Y., McTiernan C.F., Feldman A.M.;
RT   "rLERK7, rat ligand for Eph-related receptor tyrosine kinase.";
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
-!-  FUNCTION: MAY FUNCTION ACTIVELY TO STIMULATE AXON FASCICULATION.
     INDUCES COMPARTMENTALIZED SIGNALING WITHIN A CAVEOLAE-LIKE
     MEMBRANE MICRODOMAIN WHEN BOUND TO THE EXTRACELLULAR DOMAIN OF ITS
     COGNATE RECEPTOR. THIS SIGNALING EVENT REQUIRES THE ACTIVITY OF
     THE FYN TYROSINE KINASE (BY SIMILARITY).
-!-  SUBUNIT: BINDS TO THE RECEPTOR TYROSINE KINASE EPHA2, EPHA3 AND
     EPHB1 (BY SIMILARITY).
-!-  SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR IT
     IS COMPARTMENTALIZED IN DISCRETE CAVEOLAE-LIKE MEMBRANE
     MICRODOMAINS (BY SIMILARITY).
-!-  TISSUE SPECIFICITY: EXPRESSED IN BRAIN, HEART, PLACENTA AND LUNG.
-!-  SIMILARITY: BELONGS TO THE EPHRIN FAMILY.
```

```
DR   EMBL; U69279; AAC05801.1; -.
DR   ProSite; PS01299; EPHRIN; 1.
DR   PRINTS; PR01347; EPHRIN.
DR   Pfam; PF00812; Ephrin; 1.
DR   InterPro; IPR001799; Ephrin.
DR   ProDom; PD002533; Ephrin; 1.
DR   PROSITE; PS01299; EPHRIN; 1.
KW   Developmental protein; Neurogenesis; Glycoprotein; GPI-anchor; Signal.
FT   SIGNAL        1     20       POTENTIAL.
FT   CHAIN        21    228       EPHRIN-A5.
FT   CARBOHYD     37     37       N-LINKED (GLCNAC...)  (POTENTIAL).
SQ   SEQUENCE   228 AA;  26358 MW;  8559B5532D580022 CRC64;

Query Match             6.8%;  Score 167.5;  DB 1;  Length 228;
Best Local Similarity  28.8%;  Pred. No. 0.00016;
Matches   65;  Conservative  29;  Mismatches  81;  Indels  51;  Gaps  11;

Qy    33 VYWNSANKRFQAEGGYVLYPQIGDRLDLLCPRAPPGPHSSSPNYEFYKLYLVG------ 85
          ||||| ||  ||  :  ||:  ||:|||||||| ||:| |::| : :  :
Db    34 VYWNSSNPRFQ-RGDYHIDVCINDYLDVFCPHYEDSVEEDKT--ERVYLYMVNFDGYSAC 90

Qy    86 -----GAQGRRCEAPPAPNLLLTCDRPDLDLRFIKFQEYSPNLWGHEFRSHHDYYIIAT 140
               ||: |:|  : :|||||:|||:|:||:|::|::: :| |  ::
Db    91 DHTSKGFKRWECNRPHSPN------GPLKFSEKFQLFTPFSLGFFRPGRFYFISS 141

Qy   141 S------DGTEGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKFVSEMPEMERDRGAAHSLE 197
                     :||| ||  || |::  : | :: ::| :
Db   142 AIPDNGRRS-------CLK--LKVFVR----PTNSCMKTIGVRDRVPEDVNDKVENSLE 186

Qy   198 PGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAAGGLALILLGVA 243
                 |  |:|  :    :  :           ||:|| ||:|
Db   187 PADDTV---HESAEPSRG--ENAAQTPRIPSRL-----LAILLFLLA 223
```

RESULT 23
EPA1_HUMAN

```
ID   EPA1_HUMAN   STANDARD;   PRT;   205 AA.
AC   P20827;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Ephrin-A1 precursor (EPH-related receptor tyrosine kinase ligand 1)
DE   (LERK-1) (Immediate early response protein B61) (Tumor necrosis
DE   factor, alpha-induced protein 4).
GN   EPHA1 OR EFL1 OR LERK1 OR TNFAIP4.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91042512; PubMed=2233719;
RA   Holzman L.B., Marks R.M., Dixit V.M.;
RT   "A novel immediate-early response gene of endothelium is induced by
RT   cytokines and encodes a secreted protein.";
RL   Mol. Cell. Biol. 10:5830-5838(1990).
RN   [2]
RP   SUBUNIT.
RX   MEDLINE=95140419; PubMed=7838529;
RA   Kozlosky C.J., Maraskovsky E., McGrew J.T., Vandenbos T.,
RA   Teepe M., Lyman S.D., Srinivasan S., Fletcher F.A., Gayle R.B. III,
RA   Cerretti D.P., Beckmann M.P.;
RT   "Ligands for the receptor tyrosine kinases hek and elk: isolation of
RT   cDNAs encoding a family of proteins.";
RL   Oncogene 10:299-306(1995).
-!-  SUBUNIT: BINDS TO THE RECEPTOR TYROSINE KINASES EPHA2, EPHA4,
     EPHA5, EPHA6 AND EPHA7. ALSO BINDS WITH LOW AFFINITY TO EPHA1.
-!-  SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
-!-  INDUCTION: BY TNF-ALPHA AND INTERLEUKIN-1 BETA.
-!-  SIMILARITY: BELONGS TO THE EPHRIN FAMILY.
```

```
DR   EMBL; M57730; AAA58388.1; -.
DR   PIR; A36377; A36377.
DR   Genew; HGNC:3221; EFNA1.
DR   MIM; 191164; -.
DR   InterPro; IPR001799; Ephrin.
DR   Pfam; PF00812; Ephrin; 1.
DR   PRINTS; PR01347; EPHRIN.
DR   ProDom; PD002533; Ephrin; 1.
DR   PROSITE; PS01299; EPHRIN; 1.
KW   Glycoprotein; GPI-anchor; Signal.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN        18    205       EPHRIN-A1.
FT   CARBOHYD     26     26       N-LINKED (GLCNAC...)  (POTENTIAL).
SQ   SEQUENCE   205 AA;  23771 MW;  4FEFC6BF4C125IA9 CRC64;

Query Match             6.8%;  Score 166;  DB 1;  Length 205;
Best Local Similarity  27.5%;  Pred. No. 0.00018;
Matches   46;  Conservative  31;  Mismatches  74;  Indels  16;  Gaps   5;

Qy    18 LLGVILGLVSGLSLEPYVWNSANKRFQAEGGYVLYPQIGDRLDLLCPRARPPGPHSSPN-- 75
          |||   ::      : :||||:|  ||:   :  | :  || ||||||: || |  :|
Db     8 LLGLCCSLAAADRHTVFWNSSNPKFRNE-DYTIHVQLNDYDIICPHYE---DHSVADAA 63

Qy    76 YEFYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDL---DLRFTIKFQEYSPNIWGHEFRSH 132
           : :::|:  |||:|  || |  :|||:|:||  :  ||| | :|::  || |  ::
Db    64 MEQYILVLVEHEEYQLCQPQSKDQVRWQCNRPSAKHGPEKLSEKFQRFTPFTLGKEFKEG 123

Qy   133 HDYYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRK 179
          | ||: :   :    :|  :  |:|: |  |::     ::|
Db   124 HSYYYISKPIHQHEDR------CLRLKVTVSGKITHSPQAHVNPQE 163
```

RESULT 24

```
ID   EFA1_RAT       STANDARD;      PRT;   205 AA.
AC   P97553;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Ephrin-A1 precursor (EPH-related receptor tyrosine kinase ligand 1)
DE   (LERK-1) (Immediate early response protein B61).
GN   EFNA1 OR EPGL1 OR LERK1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Wistar; TISSUE=Brain;
RX   MEDLINE=95405853; PubMed=7675446;
RA   Takahashi H., Ikeda T.;
RT   "Molecular cloning and expression of rat and mouse B61 gene:
RT   implications on organogenesis.";
RL   Oncogene 11:879-883(1995).
CC   -!- SUBUNIT: BINDS TO THE RECEPTOR TYROSINE KINASES EPHA2, EPHA4,
CC       EPHA5, EPHA6 AND EPHA7. ALSO BINDS WITH LOW AFFINITY TO EPHA1 (BY
CC       SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (By
CC       similarity).
CC   -!- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; D38056; BAA07242.1; -.
DR   InterPro; IPR001799; Ephrin.
DR   Pfam; PF00812; Ephrin; 1.
DR   PRINTS; PR01347; Ephrin.
DR   ProDom; PD002533; Ephrin; 1.
DR   PROSITE; PS01299; EPHRIN; 1.
KW   Glycoprotein; GPI-anchor; Signal.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN        18    205       EPHRIN-A1.
FT   CARBOHYD     26     26       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   205 AA;  23718 MW;  CB6DAB3DB56A6EAD CRC64;
```

Query Match   6.6%; Score 162.5; DB 1; Length 205;
Best Local Similarity   25.7%; Pred. No. 0.0003;
Matches 49; Conservative 33; Mismatches 86; Indels 23; Gaps   6;

```
Qy       18 LLGVLGLVSGLSLEPYVWNSANKFQAEGGVVLYFQIGDRLDLLCPRARPPGPHSSPNYE 77
            |||: ||| :| :|: :||||||:|||||||||||||||||||||||||||||||||||
Db        8 LLGLCCSLAAVDRHIVFWNSSNPKFREE-DYTVHVQLNDYLDICPHYEDDSV-ADAAME 65

Qy       78 FYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDL---DLRFTIKFQEYSPNLWGHEFRSHHD 134
            |||: || : |:| :|| : ||| :|: :|| :   :| ||| :|:| || |:: ::||
Db        66 RYSLYMVEHQEYVTCEPQSKDQVRVKCNQPEAKHGPEKLSEKFQRFTPFILGKEFKEGHS 125

Qy      135 YYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGOSPRGAVPRKPVSEMPMERD------ 189
            |||: |||| :|:| ::|: ||||||||||||||||:|:::   |:|:|:|| |
Db       126 YYVISKPIYHQ-------------------------TQCLKLKVTVNGIITHSPHAHVNPQ---EKRLQADDPEVQ 174

Qy      190 --RGAAHSLEP 198
              :  |  ||
Db       175 VLHSIGHSAAP 185
```

RESULT 25

```
ID   EFA1_MOUSE     STANDARD;      PRT;   205 AA.
AC   P52793; P97331;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Ephrin-A1 precursor (EPH-related receptor tyrosine kinase ligand 1)
DE   (LERK-1) (Immediate early response protein B61).
GN   EFNA1 OR EPGL1 OR LERK1 OR EPL1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ICR;
RX   MEDLINE=95405853; PubMed=7675446;
RA   Takahashi H., Ikeda T.;
RT   "Molecular cloning and expression of rat and mouse B61 gene:
RT   implications on organogenesis.";
RL   Oncogene 11:879-883(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c;
RA   Morris J.C., Ciarletta A., Morris G.E., Giannotti J., Caruso A.,
RA   Hammett D.J., Finnerty H., Turner K., Wood C.R.;
RL   Submitted (MAY-1995) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97060319; PubMed=8903354;
RA   Flenniken A.M., Gale N.W., Yancopoulos G.D., Wilkinson D.G.;
RT   "Distinct and overlapping expression patterns of ligands for
RT   Eph-related receptor tyrosine kinases during mouse embryogenesis.";
RL   Dev. Biol. 179:382-401(1996).
CC   -!- SUBUNIT: BINDS TO THE RECEPTOR TYROSINE KINASES EPHA2, EPHA4,
CC       EPHA5, EPHA6 AND EPHA7. ALSO BINDS WITH LOW AFFINITY TO EPHA1.
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (By
CC       similarity).
CC   -!- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
MGD; MGI:103236; Efna1.
DR   EMBL; U90662; AAB50237.1; -.
DR   EMBL; U26188; AAA67563.1; -.
DR   EMBL; D38146; BAA07344.1; -.
DR   InterPro; IPR001799; Ephrin.
DR   Pfam; PF00812; Ephrin; 1.
DR   PRINTS; PR01347; Ephrin.
DR   ProDom; PD002533; Ephrin; 1.
DR   PROSITE; PS01299; EPHRIN; 1.
KW   Glycoprotein; GPI-anchor; Signal.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN        18    205       EPHRIN-A1.
FT   CARBOHYD     26     26       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CONFLICT     74     74       A -> T (IN REF. 1).
FT   CONFLICT     79     79       H -> Y (IN REF. 1).
FT   CONFLICT     81     81       Q -> E (IN REF. 1).
FT   CONFLICT     91     91       N -> T (IN REF. 1).
FT   CONFLICT     94     94       Q -> K (IN REF. 1).
FT   CONFLICT    112    112       R -> Q (IN REF. 1).
FT   CONFLICT    115    115       T -> S (IN REF. 1).
FT   CONFLICT    138    138       S -> T (IN REF. 1).
FT   CONFLICT    154    154       S -> N (IN REF. 1).
FT   CONFLICT    156    156       Q -> H (IN REF. 1).
FT   CONFLICT    159    159       V -> A (IN REF. 1).
FT   CONFLICT    181    181       Y -> H (IN REF. 1).
```

```
FT   CONFLICT     204    204       S -> T (IN REF. 1).
SQ   SEQUENCE     205 AA;  23802 MW;  5A8F3A6E2091E868 CRC64;

Query Match     6.6%;  Score 161;  DB 1;  Length 205;
Best Local Similarity   26.1%;  Pred. No. 0.00038;
Matches  43;  Conservative  33;  Mismatches  77;  Indels  12;  Gaps  4;

Qy    18  LLGVLGLSGLSLEPVYWNSANKRFQAEGGGYVLYFQGDRLDLLCPRARPGPHSSPNYE    77
              |||||  :| | ||:||:| ||:||:|:||| :|  |::  | |::|| || ::
Db     8  LLGLCCSLAAADRHIVFWNSSNPKFREE-DYTVHVQLNDYLDIICPHYEDDSV-ADAAME  65

Qy    78  FYKLYLVGGAQGRRCEAPPAPNLLLTCDRPDL----DLRFTIKFQEYSPNLWGHEFRSHHD  134
              :|| | :  :| || |  | ::|  |   |    |||| ||:|| |||  |:||:|
Db    66  RYTLYMVEHQEYVACQPQSKDQVRWNCNRPSAKHGPEKLSEKFQRFTPFILCKEFKEGHS  125

Qy   135  YYIIATSDGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRK  179
              ||  :   :       |: :  |           |  :| |
Db   126  YYYISKPIHQE------SQCLKLKVTVNGKITHNPQAHVANPQE  163
```

RESULT 26
EFA4_MOUSE
```
ID   EFA4_MOUSE     STANDARD;      PRT;   206 AA.
AC   O08542; O55218;
DT   01-NOV-1997 (Rel. 35, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Ephrin-A4 precursor (EPH-related receptor tyrosine kinase ligand 4)
DE   (LERK-4).
GN   EFNA4 OR EPLG4 OR LERK4 OR EPL4.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97060319; PubMed=8903354;
RA   Flenniken A.M., Gale N.W., Yancopoulos G.D., Wilkinson D.G.;
RT   "Distinct and overlapping expression patterns of ligands for
RT   Eph-related receptor tyrosine kinases during mouse embryogenesis.";
RL   Dev. Biol. 179:382-401(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129;
RX   MEDLINE=98126446; PubMed=9455306;
RA   Cerretti D.P., Nelson N.;
RT   "Characterization of the genes for mouse LERK-3/Ephrin-A3 (Epl3),
RT   mouse LERK-4/Ephrin-A4 (Epl4), and human LERK-6/Ephrin-A2 (EPLG6):
RT   conservation of intron/exon structure.";
RL   Genomics 47:131-135(1998).
-!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (By
    similarity).
-!- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.
```
```
DR   EMBL; U90663; AAB50238.1; -.
DR   EMBL; U28890; AAC39962.1; -.
DR   EMBL; U92889; AAC39962.1; JOINED.
DR   MGD; MGI:106643; Efna4.
DR   InterPro; IPR001799; Ephrin.
DR   Pfam; PF00812; Ephrin; 1.
DR   PRINTS; PR01347; EPHRIN.
DR   ProDom; PD002533; Ephrin; 1.
DR   PROSITE; PS01299; EPHRIN; 1.
KW   Glycoprotein; GPI-anchor; Signal.
FT   SIGNAL       1     25       POTENTIAL.
FT   CHAIN       26    206       EPHRIN-A4.
FT   CARBOHYD    33     33       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    98     98       N-LINKED (GLCNAC...) (POTENTIAL).
FT   SITE        41     43       CELL ATTACHMENT SITE (POTENTIAL).
FT   CONFLICT     1      4       MRLL -> MLLRLGLIYPPTRPPAPPGPLV (IN REF.
FT                               1).
SQ   SEQUENCE    206 AA;  22861 MW;  43501971DD1C6EA5 CRC64;

Query Match     6.6%;  Score 160.5;  DB 1;  Length 206;
Best Local Similarity   28.1%;  Pred. No. 0.0004;
Matches  61;  Conservative  16;  Mismatches  81;  Indels  59;  Gaps  10;

Qy    32  PVYWNSANKRFQAEGGGYVLYFQGIGDRLDLLCPRARPGPHSSPNYEFYKLYLVGGAQGRR  91
              |:||||:|:|  |||| :::  || ::::||:||:| |||||:| || | :| :|| :
Db    29  PIYWNSSNPRL-LRGDAVVELGFNDYLDIFCPHYESPGPPEGP--ETFALYMVDWSGYEA  85

Qy    92  CEAPPAPNL-LLTCDRPDL---DLRFTIKFQEYSPNLWGHEFRSHHDYYIIATSDGTREG  147
             | |  | |  :: :|  |    | ||::||:  |  | ||:||: : |||:|||::|||
Db    86  CTAEGANAFQRWNCNSMPFAPFSPVRFSEKIQRYTPFPLGFEBFLPGETYVYYISVPTPESPG  145

Qy   148  -LESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKPVSEMPMRDRGAAH-SLEPGKENLPG  205
              :: |:  | |:              ::|::|:|||          ||  |   | :|
Db   146  RCLRLQVSVCC---------------------------------KESGSSHESAHP-----VG  170

Qy   206  DPTSNAIS--RGAEGPLPPPSMPAVAGAAGGIALILII  240
              :|: :    || :| |||    | ||     |||||
Db   171  SPGESGTSGWRGGHAPSP----------------LCLLLL  194
```

RESULT 27
CA1A_MOUSE
```
ID   CA1A_MOUSE     STANDARD;      PRT;   680 AA.
AC   Q05306;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   Collagen alpha 1(X) chain precursor.
GN   COL10A1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c;
RX   MEDLINE=93143676; PubMed=8424763;
RA   Elima K., Eerola I., Rosati R., Metsaranta M., Garofalo S., Perala M.,
RA   de Crombrugghe B., Vuorio E.;
RT   "The mouse collagen X gene: complete nucleotide sequence, exon
RT   structure and expression pattern.";
RL   Biochem. J. 289;247-253(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/Sv;
RX   MEDLINE=93238750; PubMed=8477738;
RA   Kong R.Y.C., Kwan K.M., Lau E.T., Thomas J.T., Boot-Handford R.P.,
RA   Grant M.E., Cheah K.S.E.;
RT   "Intron-exon structure, alternative use of promoter and expression of
RT   the mouse collagen X gene, Col10a-1.";
RL   Eur. J. Biochem. 213:99-111(1993).
RN   [3]
RP   SEQUENCE OF 51-680 FROM N.A.
RC   STRAIN=DBA/2J;  TISSUE=Liver;
RX   MEDLINE=92267014; PubMed=1587271;
RA   Apte S.S., Seldin M.F., Hayashi M., Olsen B.R.;
RT   "Cloning of the human and mouse type X collagen genes and mapping of
RT   the mouse type X collagen gene to chromosome 10.";
RL   Eur. J. Biochem. 206:217-224(1992).
RN   [4]
RP   SEQUENCE OF 385-627 FROM N.A.
RC   STRAIN=C57BL/6;
RX   MEDLINE=99182017; PubMed=1543751;
RA   Elima K., Metsaranta M., Kallio J., Peraelae M., Eerola I.,
```

RA   Garofalo S., de Crombrugghe B., Vuorio E.;
RT   "Specific hybridization probes for mouse alpha 2(IX) and alpha 1(X)
RT   collagen mRNAs.";
RL   Biochim. Biophys. Acta 1130:78-80(1992).
CC   -!- FUNCTION: TYPE X COLLAGEN IS A PRODUCT OF HYPERTROPHIC
CC   CHONDROCYTES AND HAS BEEN LOCALIZED TO PRESUMPTIVE
CC   MINERALIZATION ZONES OF HYALINE CARTILAGE.
CC   -!- SUBUNIT: HOMOTRIMER.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC   UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC   -!- SIMILARITY: STRONG, TO ALPHA 1 AND 2 TYPE VIII COLLAGENS.
CC   -!- SIMILARITY: CONTAINS 1 C1Q DOMAIN.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; X67348; CAA47763.1; -.
DR   EMBL; X65121; CAA46237.1; -.
DR   EMBL; X63013; CAA44741.1; -.
DR   EMBL; Z21610; CAA79736.1; -.
DR   PIR; S28807; S28807.
DR   PIR; S31216; S31216.
DR   PIR; S22215; S22215.
DR   MGD; MGI:88445; Col10a1.
DR   InterPro; IPR001073; Collagen.
DR   InterPro; IPR000087; Collagen.
DR   Pfam; PF00386; C1q; 1.
DR   Pfam; PF01391; Collagen; 9.
DR   PRINTS; PR00007; COMPLEMNTC1Q.
DR   ProDom; PD000007; Collagen; 2.
DR   SMART; SM00110; C1Q; 1.
DR   PROSITE; PS01113; C1Q; 1.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Cartilage; Collagen; Signal.
FT   SIGNAL   1   18   POTENTIAL.
FT   CHAIN   19   680   COLLAGEN ALPHA 1(X) CHAIN.
FT   DOMAIN   19   56   NONHELICAL REGION (NC2).
FT   DOMAIN   57   519   TRIPLE-HELICAL REGION.
FT   DOMAIN   520   680   NONHELICAL REGION (NC1).
FT   DOMAIN   545   680   C1Q.
FT   CONFLICT   248   248   P -> L (IN REF. 3).
FT   CONFLICT   286   286   A -> S (IN REF. 2).
FT   CONFLICT   306   306   A -> S (IN REF. 3).
FT   CONFLICT   417   417   T -> S (IN REF. 3).
FT   CONFLICT   451   451   R -> K (IN REF. 4).
FT   CONFLICT   500   500   H -> R (IN REF. 3).
FT   CONFLICT   567   567   A -> C (IN REF. 3).
FT   CONFLICT   569   569   I -> H (IN REF. 3).
FT   CONFLICT   571   572   FD -> IY (IN REF. 3).
FT   CONFLICT   635   635   Q -> T (IN REF. 3).
SQ   SEQUENCE   680 AA; 66775 MW; FE984CA99AF708E2 CRC64;

Query Match   6.5%; Score 159; DB 1; Length 680;
Best Local Similarity   25.5%; Pred. No. 0.0017;
Matches 97; Conservative 30; Mismatches 108; Indels 146; Gaps 24;

Qy   1 MGPPHSGPGGV-RVGALLLLGVLGLVSGLSLEPVYWNSANKRPQAEGGYVLYPQIGDRLD 59
               ::|| ::| ||:||| |||| |::::||:::
Db   211 IGPP--GPSGVGRGENGPPGQPGI---------------KGDRGPPGEWG--------- 244

Qy   60 LLCPRARPPGPHSSPNYEFYKLYLVGGACGGRRCEAPAPNLLLTCDRPDLDLRFTIKFQE 119
                   |||| ::||| :| ||| ::| :|: ||: |
Db   245 ----PSGPPGQPGP--------GKQGR-EGIGKPGAIGSPGQPGI------------- 277

Qy   120 YSPNIWGHEFRSHHDYYIIATSDGT-----REGLESLQGGVCLITRGMKVLLRVGQSPRGGA 175
                    :| |:| | ::| :||: |:: |||
Db   278 --PGEKGHPGAPG----IAGPPGAPGFGKQGLPGLRG----ORG----------PAG-- 314

Qy   176 VPRKPVSEMPMERDRGAAHSLEPGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAGGL 235
                   ::| |||:| || ::|:: | | |: ||| | : | || :
Db   315 LPGAPGA----KGERGPAG--HPGEPGLPGSP----GNMGPQGPGPKGIPGNHGIPGAKGEI 364

Qy   236 ALILLLGVAGAGGAMCWRRRAKP---SESRHPG------------PGSFGRGGSLGLGGGGG 282
               :::|||||||| |: ||||:   ::| |||:|||||:|:|||||
Db   365 G--LVGPAGPPGA--RGARGPPGDLDGKRGYPGPGPBGLNGPKGNPGLPGLPGGKGDPGVGGTPG 419

Qy   283 M------------------------------GPREAPEPGELGIALRGGGAADPPPCPHYEKVSGDYGHPVY 323
                                              | | ::| :|::|: |
Db   420 LRGPVGPVGAKGVPGCHNGEAGPR-GEPGIPGTR---GPTGPPGVPGFPGSKGDPGNP-- 472

Qy   324 IVQDGP-------------------PQSPP 334
                   || ||::||
Db   473 -GAPGPAGIATKGLNGPTGPP 492

RESULT 28

ID   CA13_BOVIN   STANDARD;   PRT;   1049 AA.
AC   P04258;
DT   20-MAR-1987 (Rel. 04, Created)
DT   20-MAR-1987 (Rel. 04, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Collagen alpha 1(III) chain.
GN   COL3A1.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE OF 1-242.
RX   MEDLINE=80026026; PubMed=488906;
RA   Fietzek P.P., Allmann H., Rauterberg J., Henkel W., Wachter E.,
RA   Kuhn K.;
RT   "The covalent structure of calf skin type III collagen. I. The amino
RT   acid sequence of the amino terminal region of the alpha 1(III) chain
RT   (positions 1-222).";
RL   Hoppe-Seyler's Z. Physiol. Chem. 360:809-820(1979).
RN   [2]
RP   SEQUENCE OF 243-422.
RX   MEDLINE=80026027; PubMed=488907;
RA   Dewes H., Fietzek P.P., Kuhn K.;
RT   "The covalent structure of calf skin type III collagen. II. The amino
RT   acid sequence of the cyanogen bromide peptide alpha 1(III)CB1,8,10,2
RT   (positions 223-402).";
RL   Hoppe-Seyler's Z. Physiol. Chem. 360:821-832(1979).
RN   [3]
RP   SEQUENCE OF 423-571.
RX   MEDLINE=80026028; PubMed=488908;
RA   Bentz H., Fietzek P.P., Kuhn K.;
RT   "The covalent structure of calf skin type III collagen. III. The
RT   amino acid sequence of the cyanogen bromide peptide alpha 1(III)CB4
RT   (positions 403-551).";
RL   Hoppe-Seyler's Z. Physiol. Chem. 360:833-840(1979).
RN   [4]
RP   SEQUENCE OF 572-808.
RX   MEDLINE=80026029; PubMed=488909;
RA   Lang H., Glanville R.W., Fietzek P.P., Kuhn K.;
RT   "The covalent structure of calf skin type III collagen. IV. The amino
RT   acid sequence of the cyanogen bromide peptide alpha 1(III)CB5
RT   (positions 552-788).";
RL   Hoppe-Seyler's Z. Physiol. Chem. 360:841-850(1979).
RN   [5]
RP   SEQUENCE OF 809-947.
RX   MEDLINE=80026030; PubMed=488910;
RA   Dewes H., Fietzek P.P., Kuhn K.;
RT   "The covalent structure of calf skin type III collagen. V. The amino
RT   acid sequence of the cyanogen bromide peptide alpha 1(III)CB9A
RT   (position 789-927).";
RL   Hoppe-Seyler's Z. Physiol. Chem. 360:851-860(1979).
RN   [6]

RP   SEQUENCE OF 948-1049.
RX   MEDLINE=80020031; PubMed=488911;
RA   Allmann H., Fietzek P.P., Glanville R.W., Kuhn K.;
RT   "The covalent structure of calf skin type III collagen. VI. The amino
RT   acid sequence of the carboxyterminal cyanogen bromide peptide alpha
RT   1(III)CB9B (positions 928-1028)."
RL   Hoppe-Seyler's Z. Physiol. Chem. 360:861-868(1979).
RN   [2]
RL   FUNCTION: COLLAGEN TYPE III OCCURS IN MOST SOFT CONNECTIVE TISSUES
CC   ALONG WITH TYPE I COLLAGEN.
CC   -!- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(III) CHAINS. THE CHAINS ARE
CC   LINKED TO EACH OTHER BY INTERCHAIN DISULFIDE BONDS. TRIMERS ARE
CC   ALSO CROSS-LINKED VIA HYDROXYLYSINES.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC   UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
DR   PIR; A02862; CGBO7S.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR001007; VWF_C.
DR   Pfam; PF01391; Collagen; 17.
DR   ProDom; PD000007; Collagen; 1.
DR   PROSITE; PS01208; VWFC; PARTIAL.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Glycoprotein; Collagen.
FT   DOMAIN   1   14   NONHELICAL REGION (N-TERMINAL).
FT   DOMAIN   15   1040   TRIPLE-HELICAL REGION.
FT   DOMAIN   1041   1049   NONHELICAL REGION (C-TERMINAL).
FT   MOD_RES   95   95   HYDROXYLATION.
FT   MOD_RES   107   107   HYDROXYLATION.
FT   MOD_RES   119   119   HYDROXYLATION.
FT   MOD_RES   938   938   HYDROXYLATION.
FT   MOD_RES   950   950   HYDROXYLATION.
FT   CARBOHYD   107   107   O-LINKED (GAL...).
FT   CARBOHYD   950   950   O-LINKED (GAL...).
FT   DISULFID   1040   1040   INTERCHAIN.
FT   DISULFID   1041   1041   INTERCHAIN.
FT   SEQUENCE   1049 AA; 93651 MW; 8BEC33D1C66EC9A3 CRC64;

Query Match   26.5%; Score 154.5; DB 1; Length 1049;
Best Local Similarity   6.3%;
Matches   60; Conservative   9; Mismatches   74; Indels   83; Gaps   8;

Qy   171 PRGGAVPRKPVSEMPERDRGA---------------------------AHSLEPGKEN 202
              ::::::   ::|||:
Db   688 PAGGSGPAGPPGPQGVKGERGSPGGPGGPGAAGPGPGGPGSNGNDGPPGGSSGAPGKDG 747

Qy   203 LPGDPTSNAT-----------SRGAEGPLPPPSMPAVAGAAGGLALILLGVAGA 245
                                  ||||||| ||.:|:::|||
Db   748 PPGPPGSNGAPGSPGISGPKGDSGPPGERGAPGPQGPGPAGPGPLGIAG------LTGARGL 802

Qy   246 GGAMCWRRRAKP------SESHPGP----GSFGRGGSLGLGGGGGMGPREARPGE 292
              ||||||| |||| |||||||||||||.| ||
Db   803 AGPPGCMPGPARGSPGPQGIKGENGKPGPSGCQNGERGPPGPGPGGQLPGLPGLTRAGEPGRDGNPGS 862

Qy   293 LGIALRGG---------GAADPPPCPHYEKVSGDYGHPVVIVQDGPP 330
              |:|| |||||||.|
Db   863 DGLPGRDGAPGAKGDRGENGSPGAP-----GAPGHP----GPP 896

CC02_CAEEL
ID   CC02_CAEEL   STANDARD;   PRT;   301 AA.
AC   P17656;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-AUG-1990 (Rel. 15, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Cuticle collagen 2 precursor.
GN   COL-2 OR W016.7.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;

RX   MEDLINE=83050944; PubMed=7139711;
RA   Kramer J.M., Cox G.N., Hirsh D.;
RT   "Comparisons of the complete sequences of two collagen genes from
RT   Caenorhabditis elegans.";
RL   Cell 30:599-606(1982).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RX   MEDLINE=85105075; PubMed=2578467;
RA   Kramer J.M., Cox G.N., Hirsh D.;
RT   "Expression of the Caenorhabditis elegans collagen genes col-1 and
RT   col-2 is developmentally regulated.";
RL   J. Biol. Chem. 260:1945-1951(1985).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
CC   Coles L.;
CC   Submitted (DEC-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: NEMATODE CUTICLES ARE COMPOSED LARGELY OF COLLAGEN-LIKE
CC   PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A
CC   BARRIER TO PROTECT THE WORM FROM ITS ENVIRONMENT.
CC   -!- SUBUNIT: COLLAGEN POLYPEPTIDE CHAINS ARE COMPLEXED WITHIN THE
CC   CUTICLE BY DISULFIDE BONDS AND OTHER TYPES OF COVALENT CROSS-
CC   LINKS.
CC   -!- SIMILARITY: BELONGS TO THE CUTICULAR COLLAGEN FAMILY.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; J01048; AAA27990.1; -.
DR   EMBL; V00148; CAA23464.1; -.
DR   EMBL; Z68301; CAA92620.1; -.
DR   PIR; B31219; B31219.
DR   WormPep; W01B6.7; CE03759.
DR   InterPro; IPR002486; Col_cuticle_N.
DR   Pfam; PF01484; Col_cuticle_N; 1.
KW   Cuticle; Connective tissue; Repeat; Multigene family; Collagen;
KW   Signal.
FT   SIGNAL   1   37   POTENTIAL.
FT   CHAIN   38   301   CUTICLE COLLAGEN 2.
FT   DOMAIN   85   97   GLY-RICH.
FT   DOMAIN   105   134   TRIPLE-HELICAL REGION.
FT   DOMAIN   153   176   TRIPLE-HELICAL REGION.
FT   DOMAIN   183   212   TRIPLE-HELICAL REGION.
FT   DOMAIN   215   282   TRIPLE-HELICAL REGION.
SQ   SEQUENCE   301 AA; 33317E3BDAC302F9 CRC64;

Query Match   28.7%; Score 151.5; DB 1; Length 301;
Best Local Similarity   6.2%;
Matches   50; Conservative   15; Mismatches   72; Indels   37; Gaps   7;

Qy   168 GQSPRGGAVPRKPVSEMPERDRG----------AAHSLEPGKENLPGDPT--SNATSRGAE 217
              :|:.::|||: |::::| |:| :|||||||||.|||| ||||||
Db   129 GNPGKGRASAPCRPVTQPPCQPCPGGPPGPPAGPPAGPPGPPGDGNPGSPAGPSGPGPAGPP 188

Qy   218 GPLPPPSMPAVAGAAGGLALILILGVAGA---GGAMCWRRRRAKPSESRHGP-GSFGRGG 273
              |||.||||||||||||||||||||| |||||||||||||||.|| ||||||
Db   189 GPAGPAGNDGAPGAPGG-----PGEPGASEQGGP------GEPGPAGPGPGPAGPAGNDG 236

Qy   274 SLGLGGGGGMGPR----------EARPGELGIALRGGGGAADPPPPCPHYEKVSG 316
              |:|.||||||:   :|||.::|||:|||||.|| ||||||||
Db   237 APGTGPGPGPAGPGKGPGPCGAAGPGADGNPGGPGTAGKPGGEGGEKGICPKYCAIDG 290

CA34_HUMAN

```
ID   CA34_HUMAN     STANDARD;     PRT;  1670 AA.
AC   Q01955; Q9BQT2;
DT   01-OCT-1996 (Rel. 34, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Collagen alpha 3(IV) chain precursor (Goodpasture antigen).
GN   COL4A3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Kidney;
RX   MEDLINE=94364994; PubMed=8083201;
RA   Mariyama M., Leinonen A., Mochizuki T., Tryggvason K., Reeders S.T.;
RT   "Complete primary structure of the human alpha 3(IV) collagen chain.
RT   Coexpression of the alpha 3(IV) and alpha 4(IV) collagen chains in
RT   human tissues.";
RL   J. Biol. Chem. 269:23013-23017(1994).
RN   [2]
RP   REVISIONS.
RA   Leinonen A.;
RL   Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A., VARIANTS AS E-297; R-407; R-640; R-1167; E-1207;
RP   Q-1215; S-1277; T-1330; E-1334; E-1347 AND C-1661, AND VARIANTS R-43;
RP   E-162; Y-326; H-408; R-451; L-574; E-1269 AND P-1474.
RX   MEDLINE=21064696; PubMed=11134255;
RA   Heidet L., Arrondel C., Forestier L., Cohen-Solal L., Mollet G.,
RA   Gutierrez B., Stavrou C., Gubler M.C., Antignac C.;
RT   "Structure of the human type IV collagen gene COL4A3 and mutations in
RT   autosomal Alport syndrome.";
RL   J. Am. Soc. Nephrol. 12:97-106(2001).
RN   [4]
RP   SEQUENCE OF 1386-1670 FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=93015826; PubMed=1402091;
RA   Quinones S., Bernal D., Garcia-Sogo M., Elena J.M., Saus J.;
RT   "Exon/intron structure of the human alpha 3(IV) gene encompassing the
RT   Goodpasture antigen (alpha 3(IV)NC1). Identification of a potentially
RT   antigenic region at the triple helix/NC1 domain junction.";
RL   J. Biol. Chem. 267:19780-19784(1992).
RN   [5]
RP   SEQUENCE OF 1453-1670 FROM N.A.
RX   MEDLINE=91355570; PubMed=1882840;
RA   Morrison K.E., Mariyama M., Yang-Feng T.L., Reeders S.T.;
RT   "Sequence and localization of a partial cDNA encoding the human alpha
RT   3 chain of type IV collagen.";
RL   Am. J. Hum. Genet. 49:545-554(1991).
RN   [6]
RP   SEQUENCE OF 1331-1670 FROM N.A.
RX   MEDLINE=92147878; PubMed=1737849;
RA   Turner N., Mason P.J., Brown R., Fox M., Povey S., Rees A.,
RA   Pusey C.D.;
RT   "Molecular cloning of the human Goodpasture antigen demonstrates it
RT   to be the alpha 3 chain of type IV collagen.";
RL   J. Clin. Invest. 89:592-601(1992).
RN   [7]
RP   SEQUENCE OF 1644-1670 FROM N.A.
RC   TISSUE=Kidney;
RA   Ding J.;
RL   Submitted (JAN-1993) to the EMBL/GenBank/DDBJ databases.
RN   [8]
RP   SEQUENCE OF 1439-1670, AND ALTERNATIVE SPLICING.
RC   TISSUE=Kidney;
RA   Feng L., Xia Y., Wilson C.B.;
RX   MEDLINE=94124597; PubMed=8294492;
RT   "Alternative splicing of the human alpha 3(IV)
RT   collagen gene. Differential expression of mRNA transcripts that
RT   predict three protein variants with distinct carboxyl regions.";
RL   J. Biol. Chem. 269:2342-2348(1994).
RN   [9]
RP   SEQUENCE OF 1-29 FROM N.A.
RX   MEDLINE=98196554; PubMed=9537506;
RA   Momota R., Sugimoto M., Oohashi T., Kigasawa K., Yoshioka H.,
RA   Ninomiya Y., Sugimoto M., Oohashi T., Kigasawa K., Yoshioka H.,
RT   "Two genes, COL4A3 and COL4A4 coding for the human alpha3(IV) and
RT   alpha4(IV) collagen chains are arranged head-to-head on chromosome
RT   2q36.";
RL   FEBS Lett. 424:11-16(1998).
RN   [10]
RP   ALTERNATIVE SPLICING.
RX   MEDLINE=93280184; PubMed=8505332;
RA   Bernal D., Quinones S., Saus J.;
RT   "The human mRNA encoding the Goodpasture antigen is alternatively
RT   spliced.";
RL   J. Biol. Chem. 268:12090-12094(1993).
RN   [11]
RP   VARIANT PRO-1474.
RX   MEDLINE=95078827; PubMed=7987301;
RA   Lemmink H.H., Mochizuki T., van den Heuvel L.P.W.J., Schroeder C.H.,
RA   Barrientos A., Monnens L.A.H., van Oost B.A., Brunner H.G.,
RA   Reeders S.T., Smeets H.J.M.;
RT   "Mutations in the type IV collagen alpha 3 (COL4A3) gene in autosomal
RT   recessive Alport syndrome.";
RL   Hum. Mol. Genet. 3:1269-1273(1994).
CC   -!- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
CC       GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'
CC       MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENTACTIN/
CC       NIDOGEN.
CC   -!- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-
CC       ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
CC       WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
CC   -!- SUBCELLULAR LOCATION: CELL SURFACE (POTENTIAL).
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST 3 ISOFORMS; 1 (SHOWN HERE), 2/V AND
CC       3/L5; ARE PRODUCED BY ALTERNATIVE SPLICING. THEY DIFFER IN THEIR
CC       C-TERMINAL NC1 DOMAINS.
CC   -!- TISSUE SPECIFICITY: ALPHA 3 AND ALPHA 4 TYPE IV COLLAGENS ARE
CC       COLOCALIZED AND PRESENT ONLY IN BASEMENT MEMBRANES OF KIDNEY, EYE,
CC       COCHLEA, LUNG AND BRAIN.
CC   -!- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
CC       DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
CC       G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
CC       CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
CC       TRIPLE-HELICAL 7S DOMAIN.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC       UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC   -!- PTM: THE ALTERNATIVE SPLICED FORM V CONTAINS AN ADDITIONAL
CC       N-LINKED GLYCOSYLATION SITE.
CC   -!- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
CC       ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
CC       THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
CC       IV COLLAGENS.
CC   -!- PTM: Phosphorylated by the Goodpasture antigen-binding protein.
CC   -!- DISEASE: ANTIBODIES AGAINST THE NC1 DOMAIN OF ALPHA3(IV) MEDIATE
CC       THE HUMAN AUTOIMMUNE DISEASE, GOODPASTURE SYNDROME, WHICH IS
CC       CHARACTERIZED BY HEMATURIA AND PULMONARY HEMORRHAGE.
CC   -!- DISEASE: DEFECTS IN COL4A3 ARE ASSOCIATED WITH THE TYPE I
CC       AUTOSOMAL RECESSIVE FORM OF ALPORT SYNDROME, AN HEREDITARY
CC       GLOMERULONEPHROPATHY CHARACTERIZED BY PROGRESSIVE RENAL FAILURE,
CC       HEMATURIA AND DEAFNESS. THE RECESSIVE FORM OCCURS EQUALLY BETWEEN
CC       MALES AND FEMALES.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X80033; CAA56335.1; -.
DR   EMBL; AJ284487; CAC36101.1; -.
DR   EMBL; AJ284488; CAC36101.1; JOINED.
DR   EMBL; AJ284489; CAC36101.1; JOINED.
```

```
DR   EMBL; AJ288490; CAC36101.1;  JOINED.
DR   EMBL; AJ288491; CAC36101.1;  JOINED.
DR   EMBL; AJ288492; CAC36101.1;  JOINED.
DR   EMBL; AJ288493; CAC36101.1;  JOINED.
DR   EMBL; AJ288494; CAC36101.1;  JOINED.
DR   EMBL; AJ288495; CAC36101.1;  JOINED.
DR   EMBL; AJ288496; CAC36101.1;  JOINED.
DR   EMBL; AJ288497; CAC36101.1;  JOINED.
DR   EMBL; AJ288498; CAC36101.1;  JOINED.
DR   EMBL; AJ288499; CAC36101.1;  JOINED.
DR   EMBL; AJ288500; CAC36101.1;  JOINED.
DR   EMBL; AJ288501; CAC36101.1;  JOINED.
DR   EMBL; AJ288502; CAC36101.1;  JOINED.
DR   EMBL; AJ288503; CAC36101.1;  JOINED.
DR   EMBL; AJ288504; CAC36101.1;  JOINED.
DR   EMBL; AJ288505; CAC36101.1;  JOINED.
DR   EMBL; AJ288506; CAC36101.1;  JOINED.
DR   EMBL; AJ288507; CAC36101.1;  JOINED.
DR   EMBL; AJ288508; CAC36101.1;  JOINED.
DR   EMBL; AJ288509; CAC36101.1;  JOINED.
DR   EMBL; AJ288510; CAC36101.1;  JOINED.
DR   EMBL; AJ288511; CAC36101.1;  JOINED.
DR   EMBL; AJ288512; CAC36101.1;  JOINED.
DR   EMBL; AJ288513; CAC36101.1;  JOINED.
DR   EMBL; AJ288514; CAC36101.1;  JOINED.
DR   EMBL; AJ288515; CAC36101.1;  JOINED.
DR   EMBL; AJ288516; CAC36101.1;  JOINED.
DR   EMBL; AJ288517; CAC36101.1;  JOINED.
DR   EMBL; AJ288518; CAC36101.1;  JOINED.
DR   EMBL; AJ288519; CAC36101.1;  JOINED.
DR   EMBL; AJ288520; CAC36101.1;  JOINED.
DR   EMBL; AJ288521; CAC36101.1;  JOINED.
DR   EMBL; AJ288522; CAC36101.1;  JOINED.
DR   EMBL; AJ288523; CAC36101.1;  JOINED.
DR   EMBL; AJ288524; CAC36101.1;  JOINED.
DR   EMBL; AJ288525; CAC36101.1;  JOINED.
DR   EMBL; AJ288526; CAC36101.1;  JOINED.
DR   EMBL; AJ288527; CAC36101.1;  JOINED.
DR   EMBL; AJ288528; CAC36101.1;  JOINED.
DR   EMBL; AJ288529; CAC36101.1;  JOINED.
DR   EMBL; AJ288530; CAC36101.1;  JOINED.
DR   EMBL; AJ288531; CAC36101.1;  JOINED.
DR   EMBL; AJ288532; CAC36101.1;  JOINED.
DR   EMBL; AJ288533; CAC36101.1;  JOINED.
DR   EMBL; AJ288534; CAC36101.1;  JOINED.
DR   EMBL; AJ288535; CAC36101.1;  JOINED.
DR   EMBL; AJ288536; CAC36101.1;  JOINED.
DR   EMBL; AJ288537; CAC36101.1;  JOINED.
DR   EMBL; AJ288538; CAC36101.1;  JOINED.
DR   EMBL; M92993;  AAA21610.1;  -.
DR   EMBL; S57790;  AAB19637.1;  -.
DR   EMBL; M81379;  AAA51556.1;  -.
DR   EMBL; L08650;  AAA52044.1;  -.
DR   EMBL; U02519;  AAA18942.1;  -.
```

```
Db   1165 PAG----EKGETGLLRAPPGPRGNPGAQGAK--------GDRGA-----PGFFGLPG--- 1204
                  |||||||||||                     |||||         |||||||
Qy    209 SNATSRGAEGPLPPPSMPAVAGAAGGLALILLGVAGAGGAMCWRRRAKPSESRHPGP-- 266

Db   1205 -RKGAMGDAGPRGPTGIEGPPGPPGLPGAIIPGQTGNRGP---PGSRGSPSGAPGPPG 1260
                  ||:  || |  ||:|| ||||||| |  | |       |||:|||| |||||||
Qy    267 --------GSFGRGGSLG-LGGGGGMGP--REAEPGELGIALRGGGAADPPFCHYEK 313

Db   1261 SHVIGIKGDKGSMGHPGPKGPPGTAGDMGPPGRLGAPGTPGLP---GPRGDPGF-QGFPG 1316
                  ||||||| || |::||  ||  ||||||| |||  | ||||   ||||||| |||||
Qy    314 VSGDYGHPVYIVQDGP------PQSPPNI 336

Db   1317 VKGEKGNPGFLGSSIGPPGPIGPKGPPGV 1344
                  ||| | |||| || ||||||||||||||
Qy    314 VSGDYGHPVYIVQDGP-----PQSPPNI 336
```

```
RESULT 31

ID   CA1H_MOUSE                STANDARD;     PRT;   1527 AA.
AC   P39061; Q62002; Q61437;
DT   01-FEB-1995 (Rel. 31, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Collagen alpha 1(XVIII) chain precursor [Contains: Endostatin].
GN   COL18A1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A. (SHORT ISOFORM).
RC   STRAIN=BALB/c; TISSUE=Liver;
RX   MEDLINE=94245707; PubMed=8188673;
RA   Rehn M.V., Hintikka E., Pihlajaniemi T.;
RT   "Primary structure of the alpha 1 chain of mouse type XVIII collagen,
RT   partial structure of the corresponding gene, and comparison of the
RT   alpha 1(XVIII) chain with its homologue, the alpha 1(XV) collagen
RT   chain.";
RL   J. Biol. Chem. 269:13929-13935(1994).
RN   [2]
RP   SEQUENCE FROM N.A. (SHORT AND LONG ISOFORMS).
RA   Rehn M., Hintikka E., Pihlajaniemi T.;
RT   "Characterization of the mouse gene for the alpha-1 chain of type
RT   XVIII collagen (Col18A1) reveals that the three variant N-terminal
RT   polypeptide forms are transcribed from two widely separated
RT   promoters.";
RL   Submitted (NOV-1993) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 213-1140 FROM N.A. (SHORT ISOFORM).
RX   MEDLINE=94240112; PubMed=8183894;
RA   Rehn M.V., Pihlajaniemi T.;
RT   "Alpha 1(XVIII), a collagen chain with frequent interruptions in the
RT   collagenous sequence, a distinct tissue distribution, and homology
RT   with type XV collagen.";
RL   Proc. Natl. Acad. Sci. U.S.A. 91:4234-4238(1994).
RN   [4]
RP   SEQUENCE OF 240-1527 FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=94240111; PubMed=8183893;
RA   Oh S.P., Kamagata Y., Muragaki Y., Timmons S., Ooshima A., Olsen B.R.;
RT   "Isolation and sequencing of cDNAs for proteins with multiple domains
RT   of Gly-Xaa-Yaa repeats identify a distinct family of collagenous
RT   proteins.";
RL   Proc. Natl. Acad. Sci. U.S.A. 91:4229-4233(1994).
RN   [5]
RP   CHARACTERIZATION OF ENDOSTATIN, AND PARTIAL SEQUENCE.
RX   MEDLINE=97160844; PubMed=9008168;
RA   O'Reilly M.S., Boehm T., Shing Y., Fukai N., Vasios G., Lane W.S.,
RA   Flynn E., Birkhead J.R., Olsen B.R., Folkman J.;
RT   "Endostatin: an endogenous inhibitor of angiogenesis and tumor
RT   growth.";
RL   Cell 88:277-285(1997).
RN   [6]
RP   X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS) OF ENDOSTATIN.
```

```
Query Match             6.1%;  Score 148.5;  DB 1;  Length 1670;
Best Local Similarity   27.1%;  Pred. No. 0.022;
Matches  105;  Conservative  22;  Mismatches 146;  Indels 115;  Gaps 23;


Qy      1 MGPPHS-GPGGV------RVGALLLLGVLGLVSGLSLEPVYWNSANKRFQAEGGYVLYPQI 54
                                     :   |   ||  |:|        :|
Db   1020 MGMPGSKGRGTLGPPGRAGRPGLIPGIHGL-QGDKGEPGY-----------SEG---- 1061

Qy     55 GDRLDLLCPRARPPGPHSSPNYEPYKLYLVGAQGRRCE-----------------APPA 97
                                                              |||
Db   1062 -------TRPGPPGPTGDPG------LPGDMGKKGEMGQPGPPGHLGPAGPEGAPGS 1105

Qy     98 PNL--LLTCDRPDLDLRFT-IKPQEYSPNLWGHE------FRSHHDYYIIATSDGTREGLES 150
                  |:|     |:   ||         :      |:|     ||:||:|        :|:
Db   1106 PGSPGLPGKPGPGHDLGFKGIKGLLGPEGPGIRGPGPLGFPPGSPGPMGIRGDQG-RDGIPG 1164

Qy    151 LQGGVCLTRGMKVLLRVGQSPRG--GAVPRKPVSEMPMERDRGAAHSLEPGKENLPGDPT 208
                  :|: |||                                    |       |||
Db   1165 VSGDYGHPVYIVQDGP------PQSPPNI 336
```

```
RX   MEDLINE=98169382; PubMed=9501087;
RA   Hohenester E., Sasaki T., Olsen B.R., Timpl R.;
RT   "Crystal structure of the angiogenesis inhibitor endostatin at 1.5-A
RT   resolution.";
RL   EMBO J. 17:1656-1664(1998).
CC   -!- FUNCTION: ENDOSTATIN POTENTLY INHIBITS ENDOTHELIAL CELL
CC       PROLIFERATION AND ANGIOGENESIS. MAY INHIBIT ANGIOGENESIS BY
CC       BINDING TO THE HEPARAN SULPHATE PROTEOGLYCANS INVOLVED IN GROWTH
CC       FACTOR SIGNALLING.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A SHORT FORM AND A LONG FORM
CC       (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC       UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC   -!- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH
CC       INTERRUPTED HELICES (FACIT) FAMILY.
CC   -----------------------------------------------------------------------
```

```
DR   EMBL; L16898; AAA37434.1; -.
DR   EMBL; U03714; AAA20657.1; -.
DR   EMBL; U03715; AAC52901.1; -.
DR   EMBL; U34606; AAC52901.1; JOINED.
DR   EMBL; U34608; AAC52901.1; JOINED.
DR   EMBL; U34609; AAC52901.1; JOINED.
DR   EMBL; U34610; AAC52901.1; JOINED.
DR   EMBL; U03718; AAC52901.1; JOINED.
DR   EMBL; U03715; AAC52901.1; JOINED.
DR   EMBL; U34611; AAC52901.1; JOINED.
DR   EMBL; U34612; AAC52901.1; JOINED.
DR   EMBL; U34613; AAC52901.1; JOINED.
DR   EMBL; U34607; AAC52902.1; -.
DR   EMBL; U34608; AAC52902.1; JOINED.
DR   EMBL; U34609; AAC52902.1; JOINED.
DR   EMBL; U34610; AAC52902.1; JOINED.
DR   EMBL; U34611; AAC52902.1; JOINED.
DR   EMBL; U34610; AAC52902.1; JOINED.
DR   EMBL; U34611; AAC52902.1; JOINED.
DR   EMBL; U34613; AAC52902.1; JOINED.
DR   EMBL; U03716; AAC52902.1; JOINED.
DR   EMBL; U03718; AAC52902.1; JOINED.
DR   EMBL; U03716; AAC52902.1; JOINED.
DR   EMBL; U03718; AAC52902.1; JOINED.
DR   EMBL; U11636; AAC52178.1; -.
DR   EMBL; U11636; AAC52178.1; -.
DR   EMBL; L22545; AAA19787.1; -.
DR   PDB; 1KOE; 16-FEB-99.
DR   MGD; MGI:88451; Col18a1.
DR   InterPro; IPR000087; Col18a1.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR003129; TSPN.
DR   Pfam; PF01391; Collagen; 8.
DR   Pfam; PF02210; Collagen; 1.
DR   SMART; SM00282; LamG; 1.
DR   SMART; SM00210; TSPN; 1.
KW   Cell adhesion; Collagen; Glycoprotein; Signal; Alternative splicing;
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   3D-structure.
FT   SIGNAL        1     26       POTENTIAL.
FT   CHAIN        27   1527       COLLAGEN ALPHA 1 (XVIII) CHAIN.
FT   CHAIN      1344   1527       ENDOSTATIN.
FT   DOMAIN       27    538       NONHELICAL REGION 1 (NC1).
FT   DOMAIN      539    565       TRIPLE-HELICAL REGION 1 (COL1).
FT   DOMAIN      566    575       NONHELICAL REGION 2 (NC2).
FT   DOMAIN      576    649       TRIPLE-HELICAL REGION 2 (COL2).
FT   DOMAIN      650    673       NONHELICAL REGION 3 (NC3).
FT   DOMAIN      674    795       TRIPLE-HELICAL REGION 3 (COL3).
FT   DOMAIN      796    818       NONHELICAL REGION 4 (NC4).
FT   DOMAIN      819    901       TRIPLE-HELICAL REGION 4 (COL4).
FT   DOMAIN      902    915       NONHELICAL REGION 5 (NC5).
FT   DOMAIN      916    957       TRIPLE-HELICAL REGION 5 (COL5).
FT   DOMAIN      958    970       NONHELICAL REGION 6 (NC6).
FT   DOMAIN      971   1043       TRIPLE-HELICAL REGION 6 (COL6).
FT   DOMAIN     1044   1053       NONHELICAL REGION 7 (NC7).
FT   DOMAIN     1054   1086       TRIPLE-HELICAL REGION 7 (COL7).
FT   DOMAIN     1087   1098       NONHELICAL REGION 8 (NC8).
FT   DOMAIN     1099   1122       TRIPLE-HELICAL REGION 8 (COL8).
FT   DOMAIN     1123   1129       NONHELICAL REGION 9 (NC9).
FT   DOMAIN     1130   1181       TRIPLE-HELICAL REGION 9 (COL9).
FT   DOMAIN     1182   1194       NONHELICAL REGION 10 (NC10).
FT   DOMAIN     1195   1212       TRIPLE-HELICAL REGION 10 (COL10).
FT   DOMAIN     1213   1527       NONHELICAL REGION 11 (NC11).
FT   CARBOHYD    338    338       N-LINKED (GLCNAC). (POTENTIAL).
FT   CARBOHYD    700    700       N-LINKED (GLCNAC). (POTENTIAL).
FT   DISULFID   1376   1516
FT   DISULFID   1478   1508
FT   SITE       1104   1106       CELL ATTACHMENT SITE (POTENTIAL).
FT   VARSPLIC      1      1       MISSING (IN SHORT ISOFORM).
FT   VARSPLIC    213    238       AVFTQLPPFQSNLQAPLGRPSAPPDF -> MAPRWHLLDVL
FT                               TSLVLLVARVSWAE (IN SHORT ISOFORM).
FT   CONFLICT    900    900       P -> L (IN REF. 4).
FT   CONFLICT    947    947       P -> F (IN REF. 4).
FT   CONFLICT    964    964       A -> V (IN REF. 4).
FT   CONFLICT   1157   1157       A -> P (IN REF. 4).
FT   CONFLICT   1157   1157       P -> L (IN REF. 4).
FT   CONFLICT   1266   1266       P -> L (IN REF. 4).
FT   CONFLICT   1276   1276       F -> L (IN REF. 4).
FT   CONFLICT   1437   1437       L -> V (IN REF. 4).
SQ   SEQUENCE   1527 AA;  156008 MW;  9645045AF10B513 CRC64;
```

```
Query Match
Best Local Similarity  6.0%;    Score 146.5;
Matches 92; Conservative 23.7%;  Pred. No. 0.027;  DB 1;  Length 1527;
            Mismatches 107;  Indels 167;  Gaps 21;

Qy    2 GPPHSGPGGVRVGALILLLGVLG-LVSGLSLEPYVWNSANKRFQAEGGYVLYPQIGDRLDL 60
          :  :  ||| :|
Db  551 GPP---GPAGPQ------GPAGPVVQSPNSQPV-------------------------- 574

Qy   61 LCPRAR-PPGPHSSPNYEFYKLYLVGGAGGRRCE-------------------------- 93
           ||| ||   |     |
Db  575 --PGAQGPGPQGPPGPPGKD------GTPGRDGEPGDPGDEDGRPGDTGPQGPPGPGTPGDVG 624

Qy   94 ----------------APPAPNLLLTCDRPDLDLRFIKFQEYSPNLWGHEPRSHHDYYIIAT 140
                            || | ||
Db  625 PKGEKGDPGIGPRGPPGP----------PGPPGPSPRQDKLTPIDME 661

Qy  141 SDGTREGLESLQGGVCLTRGMKVLLRVGQSPRGGAVPRKP--VSEMPEMPRDR-GAAHSIE 197
             |  | ||  ||               | || |
Db  662 GSGFSGDIESLRG----------PRGPPGPPGPPGPPGVPGLPGGPPGRFGINGSVA 704

Qy  198 PGKENLPGDPTSNATSRGABGP--LP-PSMPAVGAGAGGIALILLGVYAGAGGAMCWRRR 254
          || :  |||           |   ||
Db  705 PGPAGLPGVP------GKEGPPGFPGPPGPPGPPGPPGKEGP-----PGVAGQKGSV---- 747

Qy  255 RAKPBSSRHPGP-GSFGRGGSSLGLGGGGGMGPRBEAEPGELGIALRGG--GAADPPFCPHY 311
              | |  ||
Db  748 ----GDVGIPGPKGSKGDLGPIGWPGKSGLA---GSPGPVGPPGPPGPPGPPGPFAAGF 800

Qy  312 EKVSGDYGHFVYIVQ--------DGPPQSP 333
                         :  ||  :::  |||| |
Db  801 DDMEGG-GIPLWTTARSSDGLQGPPGSP 827

RESULT 32
CAFF_RIFPA
ID   CAFF_RIFPA     STANDARD;      PRT;  1027 AA.
AC   P30754;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Fibril-forming collagen alpha chain.
OS   Riftia pachyptila (Tube worm).
OC   Eukaryota; Metazoa; Pogonophora; Vestimentifera; Axonobranchia;
OC   Riftiida; Riftiidae; Riftia.
OX   NCBI_TaxID=6426;
```

```
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=93130909; PubMed=1483468;
RA   Mann K., Gaill F., Timpl R.;
RT   "Amino-acid sequence and cell-adhesion activity of a fibril-forming
RT   collagen from the tube worm Riftia pachyptila living at deep sea
RT   hydrothermal vents.";
RL   Eur. J. Biochem. 210:839-847(1992).
RN   [2]
RP   SEQUENCE OF 8-45; 525-618 AND 810-882.
RC   TISSUE=Cuticle;
RX   MEDLINE=92015209; PubMed=1920405;
RA   Gaill F., Wiedemann H., Mann K., Timpl R., Kuhn K., Engel J.;
RT   "Molecular characterization of cuticle and interstitial collagens
RT   from worms collected at deep sea hydrothermal vents.";
RL   J. Mol. Biol. 221:209-223(1991).
CC   -!- FUNCTION: FIBRIL-FORMING COLLAGEN.
CC   -!- SUBUNIT: HOMOTRIMER.
CC   -!- PTM: ALL THE "X" ARE PROBABLY GLYCOSYLATED HYDROXYLYSINES.
DR   PIR; S22915; S22915.
DR   InterPro; IPR000087; Collagen.
DR   Pfam; PF01391; Collagen; 16.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Glycoprotein; Collagen.
FT   DOMAIN        1     12       NONHELICAL REGION (N-TERMINAL).
FT   DOMAIN       13   1023       TRIPLE-HELICAL REGION.
FT   DOMAIN     1024   1027       NONHELICAL REGION (C-TERMINAL).
FT   MOD_RES      21     21       HYDROXYLATION (PARTIAL).
FT   MOD_RES      24     24       HYDROXYLATION (PARTIAL).
FT   MOD_RES      27     27       HYDROXYLATION.
FT   MOD_RES      39     39       HYDROXYLATION.
FT   MOD_RES      53     53       HYDROXYLATION.
FT   MOD_RES      54     54       HYDROXYLATION.
FT   MOD_RES      72     72       HYDROXYLATION (PARTIAL).
FT   MOD_RES      90     90       HYDROXYLATION (PARTIAL).
FT   MOD_RES      93     93       HYDROXYLATION.
FT   MOD_RES     123    123       HYDROXYLATION.
FT   MOD_RES     128    128       HYDROXYLATION (PARTIAL).
FT   MOD_RES     150    150       HYDROXYLATION (PARTIAL).
FT   MOD_RES     161    161       HYDROXYLATION (PARTIAL).
FT   MOD_RES     162    162       HYDROXYLATION.
FT   MOD_RES     164    164       HYDROXYLATION.
FT   MOD_RES     165    165       HYDROXYLATION (PARTIAL).
FT   MOD_RES     174    174       HYDROXYLATION.
FT   MOD_RES     177    177       HYDROXYLATION.
FT   MOD_RES     180    180       HYDROXYLATION.
FT   MOD_RES     183    183       HYDROXYLATION.
FT   MOD_RES     207    207       HYDROXYLATION.
FT   MOD_RES     216    216       HYDROXYLATION.
FT   MOD_RES     219    219       HYDROXYLATION.
FT   MOD_RES     228    228       HYDROXYLATION (PARTIAL).
FT   MOD_RES     237    237       HYDROXYLATION.
FT   MOD_RES     243    243       HYDROXYLATION (PARTIAL).
FT   MOD_RES     249    249       HYDROXYLATION.
FT   MOD_RES     255    255       HYDROXYLATION.
FT   MOD_RES     273    273       HYDROXYLATION (PARTIAL).
FT   MOD_RES     276    276       HYDROXYLATION (PARTIAL).
FT   MOD_RES     285    285       HYDROXYLATION (PARTIAL).
FT   MOD_RES     291    291       HYDROXYLATION (PARTIAL).
FT   MOD_RES     303    303       HYDROXYLATION (PARTIAL).
FT   MOD_RES     306    306       HYDROXYLATION.
FT   MOD_RES     312    312       HYDROXYLATION.
FT   MOD_RES     321    321       HYDROXYLATION.
FT   MOD_RES     327    327       HYDROXYLATION.
FT   MOD_RES     339    339       HYDROXYLATION.
FT   MOD_RES     342    342       HYDROXYLATION.
FT   MOD_RES     348    348       HYDROXYLATION (PARTIAL).
FT   MOD_RES     351    351       HYDROXYLATION (PARTIAL).
FT   MOD_RES     366    366       HYDROXYLATION.
FT   MOD_RES     372    372       HYDROXYLATION.
FT   MOD_RES     375    375       HYDROXYLATION.
FT   MOD_RES     381    381       HYDROXYLATION.
FT   MOD_RES     387    387       HYDROXYLATION (PARTIAL).
FT   MOD_RES     416    416       HYDROXYLATION (PARTIAL).
FT   MOD_RES     417    417       HYDROXYLATION.
FT   MOD_RES     423    423       HYDROXYLATION.
FT   MOD_RES     429    429       HYDROXYLATION.
FT   MOD_RES     432    432       HYDROXYLATION.
FT   MOD_RES     453    453       HYDROXYLATION.
FT   MOD_RES     465    465       HYDROXYLATION.
FT   MOD_RES     483    483       HYDROXYLATION.
FT   MOD_RES     500    500       HYDROXYLATION (PARTIAL).
FT   MOD_RES     503    503       HYDROXYLATION (PARTIAL).
FT   MOD_RES     506    506       HYDROXYLATION (PARTIAL).
FT   MOD_RES     513    513       HYDROXYLATION.
FT   MOD_RES     525    525       HYDROXYLATION.
FT   MOD_RES     533    533       HYDROXYLATION (PARTIAL).
FT   MOD_RES     536    536       HYDROXYLATION (PARTIAL).
FT   MOD_RES     540    540       HYDROXYLATION.
FT   MOD_RES     546    546       HYDROXYLATION.
FT   MOD_RES     551    551       HYDROXYLATION.
FT   MOD_RES     552    552       HYDROXYLATION (PARTIAL).
FT   MOD_RES     561    561       HYDROXYLATION.
FT   MOD_RES     603    603       HYDROXYLATION.
FT   SITE        610    610       IMPERFECTION IN THE GAA REPEAT.
FT   MOD_RES     621    621       HYDROXYLATION (PARTIAL).
FT   MOD_RES     627    627       HYDROXYLATION.
FT   MOD_RES     645    645       HYDROXYLATION.
FT   MOD_RES     647    647       HYDROXYLATION.
FT   MOD_RES     648    648       HYDROXYLATION.
FT   MOD_RES     663    663       HYDROXYLATION.
FT   MOD_RES     708    708       HYDROXYLATION.
FT   MOD_RES     711    711       HYDROXYLATION.
FT   MOD_RES     714    714       HYDROXYLATION.
FT   MOD_RES     717    717       HYDROXYLATION.
FT   MOD_RES     723    723       HYDROXYLATION.
FT   MOD_RES     744    744       HYDROXYLATION (PARTIAL).
FT   MOD_RES     759    759       HYDROXYLATION.
FT   MOD_RES     773    773       HYDROXYLATION (PARTIAL).
FT   MOD_RES     774    774       HYDROXYLATION.
FT   MOD_RES     783    783       HYDROXYLATION.
FT   MOD_RES     792    792       HYDROXYLATION.
FT   MOD_RES     815    815       HYDROXYLATION (PARTIAL).
FT   MOD_RES     816    816       HYDROXYLATION.
FT   MOD_RES     843    843       HYDROXYLATION.
FT   MOD_RES     849    849       HYDROXYLATION.
FT   MOD_RES     855    855       HYDROXYLATION.
FT   MOD_RES     861    861       HYDROXYLATION.
FT   MOD_RES     867    867       HYDROXYLATION.
FT   MOD_RES     888    888       HYDROXYLATION.
FT   MOD_RES     894    894       HYDROXYLATION.
FT   MOD_RES     903    903       HYDROXYLATION.
FT   MOD_RES     915    915       HYDROXYLATION (PARTIAL).
FT   MOD_RES     933    933       HYDROXYLATION.
FT   MOD_RES     939    939       HYDROXYLATION.
FT   MOD_RES     945    945       HYDROXYLATION (PARTIAL).
FT   MOD_RES     954    954       HYDROXYLATION.
FT   MOD_RES     963    963       HYDROXYLATION.
FT   MOD_RES     966    966       HYDROXYLATION.
FT   MOD_RES     984    984       HYDROXYLATION.
FT   MOD_RES     990    990       HYDROXYLATION (PARTIAL).
FT   MOD_RES    1010   1010       HYDROXYLATION (PARTIAL).
FT   MOD_RES    1011   1011       HYDROXYLATION (PARTIAL).
FT   MOD_RES    1013   1013       HYDROXYLATION (PARTIAL).
FT   MOD_RES    1014   1014       HYDROXYLATION.
FT   MOD_RES    1016   1016       HYDROXYLATION.
FT   MOD_RES    1017   1017       HYDROXYLATION.
FT   MOD_RES    1019   1019       HYDROXYLATION.
FT   MOD_RES    1020   1020       HYDROXYLATION.
FT   VARIANT     903    903       P -> A.
FT   MOD_RES    1027   1027       HYDROXYLATION (PARTIAL).
SQ   SEQUENCE   1027 AA;  94353 MW;  9B043FA77838B8C1 CRC64;

Query Match             6.0%;  Score 146;  DB 1;  Length 1027;
  Best Local Similarity   22.2%;
  Matches   82;  Conservative   24;  Mismatches  113;  Indels  150;  Gaps   17;
```

```
Qy    46  GGYVLYPQIGDRLDLLCPRARPDPGPHSSSPNYEFYKLYLVGGAQGRRCE-----------  93
                                                                              
Db     4  GPRYIQAQVGP----IGPRG-PPGPPGSPGQQ------------GVQGLRGEPGDSGPMGPIGK  50
                                                                              
Qy    94  -APPAPNLLITCDRPDLDLRFTIKFQEYSPNLWGHEFRSHHDYYIIATSDGTREGLESLQ   152
                                                                              
Db    51  RGPPGP----------------AGIAG-----------KSGDDGRDGEPGPR  75
                                                                              
Qy   153  GGVCLT-------------RGMKVL-----LRVGQSPRGG---AVPRKPVSEM   184
                                                                              
Db    76  GGIGPMGPRGAGGMPGMPGPXSHRGPRGLSGSXGEQGKSGNQGPDGGPGPAGPSGPIGPR   135
                                                                              
Qy   185  PMERDRGAAHSLEPGKENLPGDPTSNATSRGAEGPLPPPSMPAVAGAAGGLALL-----   238
                                                                              
Db   136  GQTGERG-----RDGKSGLPG-----LRGVDGLAGPPGPGPPGPIGSTGSPGFPGTPGSK   183
                                                                              
Qy   239  ------------LLGVAGAGGAMCWRRRAKPSESRHPG-PGSFGRGGSLGLGGGGM------   283
                                                                              
Db   184  GDRGRGQSGTXGAQGLQGPVGLSGQPGVAGENGHPGMGMDGANGEPGASGHSGSGLPGPSGFP   243
                                                                              
Qy   284  GPR--------------EAEPGELGIALRGG----------GAADPPFCPHYEKVSGDYGHPVYI   324
                                                                              
Db   244  GPRGMPGTAGSPGQAGAXGDGGPTGEQGRPAGPAGPGVXGSSGPP---------GDVGAPGHA   294
                                                                              
Qy   325  VQDGPPQGSP   333
                                                                              
Db   295  GEAGKRGSP   303
```

```
RESULT 33
CA1A_BOVIN
ID   CA1A_BOVIN     STANDARD;      PRT;    674 AA.
AC   P23206;
DT   01-NOV-1991 (Rel. 20, Created)
DT   01-NOV-1991 (Rel. 20, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Collagen alpha 1(X) chain precursor.
GN   COL10A1.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Cartilage;
RX   MEDLINE=91111131; PubMed=1703407;
RA   Thomas J.T., Kwan A.P.L., Grant M.E., Boot-Handford R.P.;
RT   "Isolation of cDNAs encoding the complete sequence of bovine type X
RT   collagen. Evidence for the condensed nature of mammalian type X
RT   collagen gene.";
RL   Biochem. J. 273:141-148(1991).
CC   -!- FUNCTION: TYPE X COLLAGEN IS A PRODUCT OF HYPERTHROPHIC
CC       CHONDROCYTES AND HAS BEEN LOCALIZED TO PRESUMPTIVE
CC       MINERALIZATION ZONES OF HYALINE CARTILAGE.
CC   -!- SUBUNIT: HOMOTRIMER.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC       UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC   -!- SIMILARITY: STRONG, TO ALPHA 1 AND 2 TYPE VIII COLLAGENS.
CC   -!- SIMILARITY: CONTAINS 1 C1Q DOMAIN.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X53556; CAA37624.1; -.
DR   PIR; S13301; S13301.
DR   InterPro; IPR001073; C1q.
DR   InterPro; IPR000087; Collagen.
```

```
DR   Pfam; PF00386; C1q; 1.
DR   Pfam; PF01391; Collagen; 9.
DR   PRINTS; PR00007; COMPLEMNTC1Q.
DR   ProDom; PD000007; Collagen; 1.
DR   SMART; SM00110; C1Q; 1.
DR   PROSITE; PS01113; C1Q; 1.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Cartilage; Collagen; Signal; Glycoprotein.
FT   SIGNAL        1     18
FT   CHAIN        19    674       COLLAGEN ALPHA 1(X) CHAIN.
FT   DOMAIN       19     56       NONHELICAL REGION (NC2).
FT   DOMAIN       57    519       TRIPLE-HELICAL REGION.
FT   DOMAIN      520    674       NONHELICAL REGION (NC1).
FT   DOMAIN      539    674       C1Q.
FT   DISULFID    194    197       BY SIMILARITY.
FT   DISULFID    460    463       BY SIMILARITY.
FT   MOD_RES     460             HYDROXYLATION (BY SIMILARITY).
FT   MOD_RES     463             HYDROXYLATION (BY SIMILARITY).
FT   CARBOHYD    611             N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   674 AA;  65546 MW;  CD4CA73A03E004CA CRC64;
```

```
Query Match             5.9%;  Score 145;  DB 1;  Length 674;
Best Local Similarity   24.2%;  Pred. No. 0.014;
Matches  100;  Conservative  30;  Mismatches 125;  Indels 158;  Gaps  24;
```

```
Qy    15  ALILLGVILGLVSGLSLEPVYWNSANKRFQAEGGYVLYPQIGDRLDLLCPRA--------  65
              |||:|    | || | : ::   |  |   ||| ||||||||||||||||||        
Db     6  ALILLMSLNLVHG------VPY---TERYQTPTG-IKGPPSNTKTQFPIPYAIKGKGVSLR   56
```

```
Qy    66  ------RPPGP-------HSSPNYEFYKLYLVG-------------GAQGR-RCEAPPAPNLLIT   103
                || ||        :  |  |  |  | |            || |   | :: |        
Db    57  GEQGIPGPGPGPAGPGHPGPS---------GPPGSKPGTSGPGPQCOPGPGPGPGPS---A   104
```

```
Qy   104  CDRPDLDLRFTIKFQEYSPNLWGHEFRSHHDYYIIATSDGTREGLESLQG--------   153
              :|:|  |  | :     |  ||                    : |:| |         
Db   105  TGKPGL-------PGLPGKQ-------------------GER-GLNGPKGDIGPAGLPGP   137
```

```
Qy   154  ------GVCLTRGMKVLLRVG-QSPRGGAVPR-----KPVSEMP------MERDRGAAHSL   196
                :|:|  :| ||| :|  |::   |           |: | |      |||:| |||   
Db   138  RGPPGPPGIPGPAGISVPGKPGPGPQGPTGPEGPGPRGPPGEKGTSGVPGLNGQKGEMGHCTPC   197
```

```
Qy   197  EPGKENLPGDPTSNATSRGAEGPLPPP-------SMPAVAGAGGLALLLLGVAGAGGA   248
              | | | :|      |::        | |::|  :   |: : :| |            
Db   198  RPGERGLPG--------PQGPTGPPGPPGPPGVGKRGENGLPGQPGLKGD-----QGVPGERGA   245
```

```
Qy   249  MCWRRRRAKPSESRHP----GPGSFGRGGSLGLGGGGGMGPREAXPGELGIA--------   296
              ::  |  | ::   |     ||| |     ::                            
Db   246  ----AGPSGPQGPPGEQGPEGIGKPGAPGIPQGPGPIGPMGKXQPGAPGTAGLPGAPGF   298
```

```
Qy   297  ------LRG-GGAADPPFCPHYEKVSGDYGHPVVIVQDGP------PQSPPNI   336
                |: | |   ::| |   | | | :|            ||||| |       
Db   299  GKPGLPGLKGQGRGPVGLPGSFGAKGEQGPAGHPGBAGLPGPSGNMGPQGPKGI   351
```

```
RESULT 34
CA1H_HUMAN
ID   CA1H_HUMAN     STANDARD;      PRT;   1516 AA.
AC   P39060; Q9Y6Q8; Q9Y6Q7; Q9UK38;
DT   01-FEB-1995 (Rel. 31, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Collagen alpha 1(XVIII) chain precursor [Contains: Endostatin].
GN   COL18A1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98164096; PubMed=9503365;
RA   Saarela J., Ylikarppa R., Rehn M., Purmonen S., Pihlajaniemi T.;
RT   "Complete primary structure of two variant forms of human type XVIII
RT   collagen and tissue-specific differences in the expression of the
RT   corresponding transcripts.";
```

```
RL   Matrix Biol. 16:319-328(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20289799; PubMed=10830953;
RA   Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA   Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
RA   Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
RA   Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA   Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA   Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA   Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA   Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
RA   Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H.,
RA   Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
RA   Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA   Lehrach H., Reinhardt R., Yaspo M.-L.;
RL   "The DNA sequence of human chromosome 21.";
RL   Nature 405:311-319(2000).
RN   [3]
RP   SEQUENCE OF 834-1516 FROM N.A.
RC   TISSUE=placenta;
RA   Zhi-Yong H., Biao L., Wei-Jie Z., Xiang-Fu W.;
RL   "Cloning and expression of human endostatin gene in Escherichia
RL   coli.";
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   INVOLVEMENT IN KNOBLOCH SYNDROME.
RX   MEDLINE=20400145; PubMed=10942434;
RA   Sertie A.L., Sossi V., Camargo A.A., Zatz M., Brahe C.,
RA   Passos-Bueno M.R.;
RL   "Collagen XVIII, containing an endogenous inhibitor of angiogenesis
RL   and tumor growth, plays a critical role in the maintenance of retinal
RL   structure and in neural tube closure.";
RL   Hum. Mol. Genet. 9:2051-2058(2000).
RN   [6]
RP   VARIANT ASN-1437.
RX   MEDLINE=21518361; PubMed=11606364;
RA   Iughetti P., Suzuki O., Godoi P.H., Alves V.A., Sertie A.L.,
RA   Zorick T., Soares F., Camargo A., Moreira E.S., di Loreto C.,
RA   Moreira-Filho C.A., Simpson A., Oliva G., Passos-Bueno M.R.;
RL   "A polymorphism in endostatin, an angiogenesis inhibitor, predisposes
RL   for the development of prostatic adenocarcinoma.";
RL   Cancer Res. 61:7375-7378(2001).

CC   -!- FUNCTION: COL18A PROBABLY PLAYS A MAJOR ROLE IN DETERMINING THE
CC       RETINAL STRUCTURE AS WELL AS IN THE CLOSURE OF THE NEURAL TUBE.
CC   -!- FUNCTION: ENDOSTATIN POTENTLY INHIBITS ENDOTHELIAL CELL
CC       PROLIFERATION AND ANGIOGENESIS. MAY INHIBIT ANGIOGENESIS BY
CC       BINDING TO THE HEPARAN SULPHATE PROTEOGLYCANS INVOLVED IN GROWTH
CC       FACTOR SIGNALING.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A SHORT FORM/NC1-303 AND A LONG
CC       FORM/NC-493 (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: PRESENT IN MULTIPLE ORGANS WITH HIGHEST LEVELS
CC       IN LIVER, LUNG AND KIDNEY.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC       UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC   -!- POLYMORPHISM: There is an association between a polymorphism in
CC       position 1437 and prostate cancer. Heterozygous Asn-1437
CC       individuals have a 2.5 times increased chance of developing
CC       prostate cancer as compared with homozygous Asp-1437 individuals.
CC   -!- DISEASE: Defects in COL18A1 are a cause of Knobloch syndrome (KS
CC       or KNO); an autosomal recessive disorder defined by the occurrence
CC       of high myopia, vitreoretinal degeneration with retinal
CC       detachment, macular abnormalities and occipital encephalocele.
CC   -!- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH
CC       INTERRUPTED HELICES (FACIT) FAMILY.
```

```
DR   EMBL; AF018081; AAC39658.1; -.
DR   EMBL; AF018082; AAC39659.1; -.
DR   EMBL; AL163302; CAB90482.1; -.
DR   EMBL; L22548; AAA51864.1; -.
DR   EMBL; AF184060; AAF01310.1; ALT_INIT.
DR   HSSP; P39061; 1KOE.
DR   GlycoSuiteDB; P39060; -.
DR   Genew; HGNC:2195; COL18A1.
DR   MIM; 120328; -.
DR   MIM; 267750; -.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR003129; TSPN.
DR   Pfam; PF01391; Collagen; 7.
DR   Pfam; PF02210; LamG; 1.
DR   Pfam; PF00220; TSPN; 1.
DR   SMART; SM00282; LamG; 1.
DR   SMART; SM00210; TSPN; 1.
KW   Extracellular matrix; Collagen; Connective tissue; Repeat; Hydroxylation;
KW   Cell adhesion; Collagen; Glycoprotein; Signal; Alternative splicing;
KW   Polymorphism.
FT   SIGNAL          1     23       POTENTIAL.
FT   CHAIN          24   1516       COLLAGEN ALPHA 1 (XVIII) CHAIN.
FT   CHAIN        1334   1516       ENDOSTATIN.
FT   DOMAIN         24    516       NONHELICAL REGION 1 (NC1).
FT   DOMAIN        517    550       TRIPLE-HELICAL REGION 1 (COL1).
FT   DOMAIN        551    560       NONHELICAL REGION 2 (NC2).
FT   DOMAIN        561    640       TRIPLE-HELICAL REGION 2 (COL2).
FT   DOMAIN        641    664       NONHELICAL REGION 3 (NC3).
FT   DOMAIN        665    786       TRIPLE-HELICAL REGION 3 (COL3).
FT   DOMAIN        787    809       NONHELICAL REGION 4 (NC4).
FT   DOMAIN        810    892       TRIPLE-HELICAL REGION 4 (COL4).
FT   DOMAIN        893    906       NONHELICAL REGION 5 (NC5).
FT   DOMAIN        907    948       TRIPLE-HELICAL REGION 5 (COL5).
FT   DOMAIN        949    961       NONHELICAL REGION 6 (NC6).
FT   DOMAIN        962   1034       TRIPLE-HELICAL REGION 6 (COL6).
FT   DOMAIN       1035   1044       NONHELICAL REGION 7 (NC7).
FT   DOMAIN       1045   1077       TRIPLE-HELICAL REGION 7 (COL7).
FT   DOMAIN       1078   1089       NONHELICAL REGION 8 (NC8).
FT   DOMAIN       1090   1111       TRIPLE-HELICAL REGION 8 (COL8).
FT   DOMAIN       1112   1118       NONHELICAL REGION 9 (NC9).
FT   DOMAIN       1119   1173       TRIPLE-HELICAL REGION 9 (COL9).
FT   DOMAIN       1174   1186       NONHELICAL REGION 10 (NC10).
FT   DOMAIN       1187   1204       TRIPLE-HELICAL REGION 10 (COL10).
FT   DOMAIN       1205   1516       NONHELICAL REGION 11 (NC11).
FT   CARBOHYD       68     68       N-LINKED (GLCNAC...). (POTENTIAL).
FT   CARBOHYD      129    129       N-LINKED (GLCNAC...). (POTENTIAL).
FT   CARBOHYD      164    164       N-LINKED (GLCNAC...). (POTENTIAL).
FT   CARBOHYD      691    691       N-LINKED (GLCNAC...). (POTENTIAL).
FT   CARBOHYD     1329   1329       O-LINKED (GALNAC...).
FT                                  /FTId=CAR_000150.
FT   DISULFID     1366   1506       BY SIMILARITY.
FT   DISULFID     1468   1498       BY SIMILARITY.
FT   SITE         1095   1097       CELL ATTACHMENT SITE (POTENTIAL).
FT   VARSPLIC        1    180       MISSING (IN SHORT ISOFORM).
FT   VARSPLIC      181    215       HTTEAGTLPAPPSPPSPLGRPWAPLTGPSVPPPSS -> MA
FT                                  PRCPWPWPRRRRLIDVLAPLVILLLGVRAASAEP (IN
FT                                  SHORT ISOFORM).
FT   VARIANT      1437   1437       D -> N.
FT                                  /FTId=VAR_012709.
FT   CONFLICT      428    428       F -> S (IN REF. 2).
FT   CONFLICT      841    841       I -> V (IN REF. 2).
FT   CONFLICT      877    877       V -> L (IN REF. 3).
FT   CONFLICT      886    886       P -> R (IN REF. 3).
```